人类 符号简史

[美]约瑟夫·马祖尔 著

洪万生 洪赞天 英家铭 黄俊玮 黄美伦 郑宜瑾 译

接力出版社 Publishing House | 全国百佳图书出版单位 Top 100 Publishing Houses in China

关于本书

符号是怎么来的?

数学符号是怎么来的?

在数字和符号诞生之前，人类是怎样进行数的记录和计算的?

这个全人类都懂的表达方式是如何一步一步演变成今天的样子的?

这些看似平凡的符号为我们的世界带来了何种天翻地覆的变化?

从数千年前的美索不达米亚讲到17世纪的科学革命，从丝路讲到波斯御道，从中国讲到西方，《人类符号简史》叙述了数学符号系统发展背后引人入胜的故事，详细说明了符号在人类文明起源之初的有趣形态，揭示了符号演化过程中人类思维的神奇转变。

全书以符号和数字这种独特的脉络来解读世界史，由原始人类最初的计算需求开始，逐一向读者解读了符号从无到有的规律成因，讲述了符号发展背后的历史故事，让我们审视人类的思维是如何由虚到实、化繁为简的，是一部不可多得的有关符号及人类历史的作品。

人类
符号简史

RENLEI FUHAO JIANSHI

[美]约瑟夫·马祖尔　著

洪万生　洪赞天　英家铭　黄俊玮　黄美伦　郑宜瑾　译

接力出版社
Publishing House

桂图登字：20-2016-304

图书在版编目（CIP）数据

人类符号简史/（美）约瑟夫·马祖尔著；洪万生等译. ——南宁 ： 接力出版社， 2018.5

书名原文： Enlightening Symbols： A Short History of Mathematical Notation and Its Hidden Powers

ISBN 978-7-5448-5222-7

Ⅰ.①人… Ⅱ. ①约… ②洪… Ⅲ.①社会科学－通俗读物 Ⅳ. ①C49

中国版本图书馆CIP数据核字（2017）第302624号

责任编辑：张慧芳 文字编辑：刘盛楠 美术编辑：许继云 装帧设计：许继云

责任校对：刘艳慧 高雅 责任监印：刘冬 版权联络：王燕超

社长：黄俭 总编辑：白冰

出版发行：接力出版社 社址：广西南宁市园湖南路9号 邮编：530022

电话：010－65546561（发行部） 传真：010－65545210（发行部）

http：//www.jielibj.com E－mail：jieli@jielibook.com

经销：新华书店 印制：北京鑫丰华彩印有限公司

开本：710毫米×1000毫米 1/16 印张：15.75 字数：255千字

版次：2018年5月第1版 印次：2018年5月第1次印刷

印数：00001—13000册 定价：42.00元

质量服务承诺：如发现缺页、错页、倒装等印装质量问题，可直接向本社调换。

服务电话：010－65545440

文前辅文

献给我的大哥巴瑞

他从0开始教导我

导言

一位数学家、一位音乐家和一位心理学家走进一家酒吧……

几年前，我还压根儿没想过自己会写一本关于符号史的书，那时我与一些同事在科莫湖边贝拉焦村的一家小酒吧，曾有过一段对话。那位心理学家声称，符号在人类发展出语言之前早已存在多时，而这些符号来源于人类最基本且原始的思想。那位音乐家则指出，现代乐谱主要源于生活在第一个千禧年之交的本笃会修士吉多·阿雷佐（Guido d’Arezzo），但一种更原始的符号记谱形式几乎可追溯至腓尼基人的手稿。而我，就是那位数学家，我接下来说的事让我的朋友们大吃一惊。我告诉他们，除了数字之外，数学符号——甚至代数方程式——都是近代的发明，几乎所有数学式在15世纪末之前都是以文字（vhetorical）表述的。

“什么？”心理学家大吼说，“那乘法运算呢？你是要告诉我们没有用来表示‘相乘’的符号？”

“16世纪之前没有……也许17世纪之前都没有。”

“那么等式呢？‘等于’符号是何时出现的？”音乐家问道。

“不早于……16世纪。”

“但是欧几里得无疑使用了‘加’的符号。”心理学家说，“那毕达哥拉斯定理呢？这个定理涉及了直角三角形的边长平方相加。”

“不……12世纪之前没有表示‘加’的符号！”

当我们品味着昂贵的巴罗洛红酒时，现场陷入一阵沉默。

后来证明，我的说法并不正确。更久远之前，早在公元前18世纪，埃及人便使用了表示加和减的象形文字，以人们靠近或远离的图

形，分别代表数量的加或减。而不时地，数学著作的作者大胆利用符号来作为表达的媒介。因此，从许多例证可以看出，他们尝试以图形记号来表示文字甚至整个短语。4世纪，巴赫沙里手稿（*Bakhshâlî* manuscript）中用看起来像现代加号的符号来记录负数。3世纪，亚历山大的丢番图（Diophantus of Alexandria）使用一个希腊字母来表示未知数，并利用类似朝上的箭头符号来代表减。7世纪，印度数学家婆罗门笈多（Brahmagupta）使用小黑圆点，代表我们现在称作“零”的这个新数字。到了15世纪下半叶，现代的符号才开始羞怯地进入数学的世界。当然，长久以来，人类用以表示整数的符号一直存在。

在小酒吧那一晚，我没意识到自己估算符号使用的时间应该再早几个世纪。可以确定的是，丢番图在3世纪已用了一些他自己的表示方式。然而，12世纪之前，符号并未在符号化的层次上进行运算，也即方程式的运算不是纯符号式的。或许我该宣称，正确的说法是，在16世纪之前**大部分**数学式都是以文字表述的，好让大家更加惊讶。

自从那次谈话之后，我发现绝大多数人对于16世纪之前的数学记法不是真正的符号这件事，感到非常惊奇。我们也想知道，以符号的形式来讨论代数，有什么样的好处？又有什么不足呢？

追溯符号的根源，可知它们是一种借由从事物外观或信息传递中抽象出来的模式与结构，来理解、认识与创造意义的手段。

symbol（符号）这个词来自希腊文里代表token（象征）或token of identity（身份的象征）之义的词，它结合了两个词根：*sum*（一起）和动词*ballo*（丢掷）。对“符号”一词较宽松的诠释是“放在一起”。它的词源来自一种古老的证明方式，证明某人身份或某人与他人之间的关系。一根木棒或骨头被劈成两半，有关联的两人各取其半。为了核证这个关系，这两半必须完美地契合。

再从更深的层次来看，“符号”一词意味着，当熟悉的事物与不熟悉的事物被放在一起时，会创造出某种新事物。换句话说，当一个无意识的想法与有意识的想法契合时，新的意义浮现出来。更确切地说，符号是连接有意识与无意识的想法时所得出的意义。

数学符号真能达到这样的目的吗？它们真的必然满足上述关系吗？或许符号与记法之间存在一种差异。记法来自速记，让词语简略。如果将符号视作为我们提供潜意识思考的记法，想想“＋”的情

况。这只是一个记法，起初源自拉丁文*et*的速记。是的，它来自*et*中的t。1489年，我们在约翰内斯·威德曼（Johannes Widmann）的著作《各种职业中快速且工整的计算》（*Behende und hubsche Rechenung auff allen Kauffmanschafft*）中发现这个记法。它指一种数学运算，如同and这个词。

“＋”被用于诸如2＋3＝5的算术表达中，仅告诉我们2与3之和记作5。但在代数表达里，例如$x^2+2xy+y^2$，它的意思不单是“x^2与$2xy$与y^2”。数学家将“＋”视为形成完全平方式$(x+y)^2$的黏合剂。现在可以确定的是，数学家同样将and视为一种黏合剂。或许要花一点时间才能看出上述完全平方式，但当我们注视着某物时，心里知道它有另一种更有用的形式，熟悉的符号总是为我们提供有用的关联。

一种力求纯正的方法，能够区分出符号表示与简单记法之间的差异。我抱有一种更宽容的个人观点，数字与所有非文字形式的操作性记法是不同的，但它们同样被视为符号，因为它们代表与它们本身不相似的事物。

再读一次2＋3＝5这个算术表达。在数学上，这是一个完美的句子，有名词、连词和动词。只需一秒钟，你便能读懂它并继续往下读了。即便没有察觉到自己的事实查核过程，你仍基于许多理由相信它是对的，当你还是小孩子时就被告知这件事，最终在经年累月接收了大量确凿的证据后，无须再有意识地彻底搜寻你头脑里贮存确定事实的图书馆，便能**知晓**它是正确的。

然而，对于符号的使用技术，作家与数学家之间的差异显而易见。作家为了煽动情绪或利用个人生命旅程中所领悟的种种意义来营造深入人心的情境，会自由自在地使用符号，即使使用方式与生活经验矛盾；相反，数学家除了归谬法的论证模式必须通过导出矛盾来建立整个证明体系之外，不能构成矛盾。数学符号具有明确的基本目的：为了便于理解，严谨地包装复杂的信息。

相较于数学家，作家拥有更多自由。文学上使用的符号可能受到神话和文化的羁绊，但它们以许多不同的方式被使用。艾米莉·狄金森（Emily Dickinson）在她的诗作《一个瘦长的家伙在草地》（*A Narrow Fellow in the Grass*）中，未曾使用“蛇”这个词，以避免这个词直接连接到邪恶、鬼祟和危险，尽管蕴意是一样的。约瑟夫·康拉德（Joseph Conrad）在他的著作《黑暗之心》（*Heart of*

Darkness）中，将刚果河描述为“一条伸开身子的巨蛇，将头探进大海”，唤起了关于狡猾和鬼祟的言外之意。一个作家也可能无意中使用“蛇”这个词，却绝非意指某个事物属意料之外、诡诈或危险。它可能只是一种描述方式，正如“河流像一条蛇般蜿蜒”的描写手法一样。作家也许是试图唤起一种与其文化内涵无关的意象。总是使用比喻手法来表达其实很难——或许是不可能的。

数学家会使用一种叫作“蛇引理”（snake lemma）的引理（一个小定理，用来证明主要定理的垫脚石），它涉及一种被称为“蛇图”（snake diagram）的图形——它并非意指其中存在任何邪恶、狡猾或危险的事物，而是指图形的样子看起来就像一条蛇，这同样只是一种形象的描述。

人造的数学符号，与音乐旋律中可变的、带有情绪的符号，或者诗歌中隐喻的符号，有所不同。然而，有些符号仍容易唤起潜意识中清晰聚焦的知觉和连接。符号也可能传递隐喻的思想，能够凭借相似（similarity）、类比（analogy）和貌似（resemblance）来传达意义，如同纸上的文字一样。

阅读代数式时，富有经验的数学头脑可以在极短的神经传导时间里，跨越广大无边的连接。

以每个儿童都学过的符号 π 为例。作为一个符号，它是思想的一种感官表达，这些思想通过关联而唤醒了相关暗示。在定义上，它指一个特定的比，即圆的周长除以其直径。作为一个数字，它大约等于3.14159。它可以化身为许多不同的形式。举例来说，它可以用无穷数列来呈现：

$$\pi = \frac{4}{1} - \frac{4}{3} + \frac{4}{5} - \frac{4}{7} + \frac{4}{9} \cdots$$

或是无穷乘积：

$$\pi = 2 \cdot \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdot \frac{8}{7} \cdot \frac{8}{9} \cdots$$

或是无穷连分数：

$$\pi = \cfrac{4}{1 + \cfrac{1^2}{3 + \cfrac{2^2}{5 + \cfrac{3^2}{7 + \cdots}}}}$$

它经常出现在分析学和数值计算领域。当人们在一个方程式中发现 π，机智的读者会自动想到某种与圆有关的事物潜藏在其后。因此，这个符号（当然是指现代的形式）不会愚弄那个早已熟悉它的各种伪装的数学家，因为早已了解这个符号，头脑中会下意识地浮现出它的意义。

下面是 π 的另一种伪装：设想一条河流，它受平缓坡度的影响，在均匀的易受侵蚀的沙地上流淌。理论上可预测，随着时间的流逝，

河流的真正长度，除以起始点与终点之间的直线距离，将会趋近于 π。如果你猜想这与圆有关，你猜对了。

物理学家尤金·维格纳（Eugene Wigner）在他著名的文章《数学在自然科学中不合理的有效性》（*The Unreasonable Effectiveness of Mathematics in the Natural Sciences*）里，讲述了一个生动的故事：一个统计学家尝试解释一本使用高斯分布来研究人口趋势的再版书中各个符号的意义。“那这里这个符号是什么？”一位朋友问道。

“哦！”统计学家说，“那是 π。”

“π是什么？”

“圆周长与圆直径的比。”

“嗯，可是人口确定与圆周长无关。”

维格纳讲这个故事的目的是要告诉我们，数学概念会在令人惊奇的意外情况下出现，就像河流长度和人口趋势的例子。当然，他更关心的是了解数学与物质世界那些意想不到的连接背后的原因，但他的故事也点出问题，也就是为什么这类纯数学世界里的概念会以意料之外的方式现身？

在欧几里得的《几何原本》里，符号 π不具意义（不过是古希腊字母表中的第十六个字母），即使《几何原本》包含一个不易证明的事实：任意两个圆的面积比是它们的直径上的正方形面积之比。[1]希腊数学思想独一无二的特质，在于确信这世界存在可被证明的永恒真理：任何一个圆都会被自身任一条直径等分；任何三角形的内角和永远是一个相同的常数；三维空间中恰好存在五个正多面体。在《几何原本》第二卷命题4中，欧几里得告诉我们如何证明今日觉得简单的代数恒等式，例如 $(a+b)^2=a^2+b^2+2ab$，但是在他的证明里看不到任何表示乘幂的代数符号（那些放在右上角的小数字，说明一个数字自乘多少次），在他的命题和证明里也看不到加号，这一方面是因为他的陈述和证明是几何式的，另一方面是因为他完全以叙事形式来陈述命题和证明。

亚历山大的丢番图比欧几里得晚出生五百多年。他在巨作《算术》（*Arithmetica*）中提出两个未知数的特别线性方程如$x+y=100$及$x-y=40$某种接近于代数的解法。他并非借助符号来解答，而是使用简字化的记法（syncopated notation），也就是当时相当常见的做法：省略词中间的字母。所以他的著作离不开言辞解说。[2]那是脱离以日常语言来表现数学所跨出的第一步。

没有符号，研究所有数学仍是可行的。一般而言，法律条款中不包含诸如appurtenance（从属权）、aforesaid（前述的）、behoove（理应）等法律措辞之外的符号——除了用于法律文件，这些词只有少数人会想到要使用。无论是基于传统，或者有意安排，法律都不依赖以符号来达到精确性。自然语言中的文字，诸如英文或拉丁文，可以表达严格的意义，但几乎无法达到符号代数那种无懈可击的精确性。相反，成文法极为依赖意图，我们可以想见那些熟悉法律的聪明人会发现其中的漏洞。

想象一下，如果今天仍然完全是以文字表述，没有设计精巧的大量符号，数学会是什么样子？以阿尔-花拉子密（al-Khwārizmī）所著《代数》一书中的片段为例，甚至数字也以文字来表达：

> 如果一个人问你这样的问题：“我把十分成两部分，并将其中一部分乘上另一部分，所得结果为二十一”，那么你知道其中一部分为某数，而另一部分为十减某数。

我们会把这个问题简写为：$x(10-x)=21$。

如同阿尔-花拉子密所写的解答，当中用到的语言对该问题而言是明确的。在这段话背后，可能隐含了某种惯常的程序或某种计算法则，但它的确需要花一点力气才能看出来，因为阿尔-花拉子密的《代数》不是那个时代特别具有代表性的数学。

私人未公开的文件或许有所不同。思考过程和潦草的解答可能是在草稿上进行，如同今日数学家的做法。我无从确切得知，但我猜想最早是在某种沙板上探究解答，当中使用了某种个人的记法，之后以文字表达的方式组合，以说明文本内容。

创造力丰富的6世纪印度数学家和天文学家阿耶波多（Aryabhatta），用字母来代表未知数。而7世纪的印度数学家及天文

学家婆罗门笈多——顺带一提，他是第一位将零视为数字的作者——使用缩写来代表出现在特定问题中每一个未知数的平方与平方根。阿耶波多和婆罗门笈多都以韵文书写，因此不管他们使用什么符号体系都必须符合韵律。当读者看见一个小圆点时，他们所读的是代表小圆点的那个字。这让符号的使用受到限制。负数用一个小圆点来区分，分数的写法与今日相同，只是分子与分母中间没有横杠。

即使到了16世纪初期，欧洲的数学著作本质上仍是以文字表述的，尽管一些国家几百年来无疑经常使用缩写的文字。缩写变得简略，而到了下一世纪，通过弗朗索瓦·韦达（François Viète）、罗伯特·雷科德（Robert Recorde）、西蒙·史蒂文（Simon Stevin）及最终笛卡儿的书写，那些缩写变得非常简洁，所有与这些缩写的源头曾经显而易见的连接，从此消失无踪。

在数学里，一段文字表述的符号形式不只是便利的速记而已。首先，它不专属于任何特定的语言，世界上几乎所有语言都使用那些相同的记法，尽管书写形式可能各有不同。其次，且或许最重要的是，符号帮助思维超越那些以自然语言所写的文字伴随的模棱两可和误解。符号使得思维可以将特殊表达提升至一般化的形式。举例来说，文字表达句“从一未知数的平方，减去该未知数的两倍，再加一”可以写成x^2-2x+1。这个符号式可以提示一种集合式的表达概念，就像我们可能从x^2-2x+1的个别特性，导出一般二次式ax^2+bx+c。我们仅将x^2-2x+1视为一个类别（species）的代表。[3]

到了17世纪之初笛卡儿的时代，下述文字表述已几乎完全以现代的符号形式书写了：

> 一未知量与一个数的和的平方，等于该未知量与该数之平方和，再加上该未知量与该数之积的两倍。

当时是以符号∝来代表相等的：

$$(x+a)^2 \propto x^2+a^2+2ax$$

符号最终使得代数从文字的非形式性解放出来。

随着这一切的发展，某种东西遗失了。我们传达现代数学时主要是通过成套的符号，也就是由符号所标注的信息公文包。而那些公文包往往如同俄罗斯套娃，一个套着一个的公文包集合，每一个都取决于下一个更小的公文包的符号。

有一个关于说笑话者的老笑话：一个家伙走进酒吧，听到几个老家伙围坐在一起讲笑话。其中一个人大喊：“五十七！”而其他人放声大笑。另一个人大叫：“八十二！”大家又都笑了起来。

于是这个家伙问酒保：“发生了什么事？”

酒保答道：“噢，他们已经在这里一起厮混讲笑话很久了，把他们所有的笑话以数字来编目。当他们要说笑话时只需要喊出那个数字。这样比较省时。”

这个刚到的家伙说：“真聪明！我来试试。”

于是这个家伙转向那群年长者大喊：“二十二！”

所有人只是看着他，没有人笑。

他尴尬地坐下，问那个酒保：“为什么没有人笑？”酒保说：“嗯，年轻人，你只是没有说对方式……”

数学家通常通过一系列符号信息和编码来沟通，这些符号对于没有钥匙打开装满内容的那些公文包的新手而言，是难以理解的。那些比人类曾创造的所有自然语言更难学习的记号、标志和符号所造成的困难，让数学家们变得非常小众。

为了帮助理解，数学家多半在言谈中放宽他们无懈可击的论证，牺牲绝对严密的证明。他们仰赖所谓的“口语的灵活性”，一种通过共通的专业基础知识与独立于文化背景的经验，来彼此理解的方式。

然而，即使运用口语的灵活性，绝对证明之外的某种东西还是消失了。数学，甚至应用数学、物理和化学，都可以在仅有图形符号且没有任何可想象得到的实物做参照的情况下进行研究。因此，物理学家所用的文字说明与数学家的文字说明之间的差异，是一种概念化的差异。

这或许恰好可以说明为什么物理学家比较容易与大众沟通，他们能够为我们说明这个世界上的“玩意儿”。他们所讲的玩意儿可能是银河、撞球、原子、物质的基本粒子和弦，但即使人们察觉不到那些弦，它们存在于十维空间且小于10^{-35}米，却可以把它们想象成一种玩意儿。甚至电场和磁场也可以被想象成某种玩意儿。当物理学家撰写一本大众读物时，他们一开始便占有优势，他们知道每一位读者都曾体验过他们用语言所描述的事物，因为连他们提到的大部分无穷小的物体都是可以想象的“东西”。

数学家所使用的基本要素是某种更难以捉摸的东西。表示一个特定数字的符号*N*，不只是一种便利的记法。现今它在意识中代表一种与这个世界少有关联的事物——换言之，*N*是意识中的一个“存在”（being），而非这个世界中一个确定的“存在”。所以，这个非实体的事物具有一种认知层面上的本体论。现代理解一个数字——例如三——的思维过程，就像理解任何抽象事物一样，是一个爬梯的过程，都是从人类经验中确定数量的事物开始，逐渐超越到一般化事物：田野中的三只羊，三只羊，三个生物，三个东西……一路爬升到“三”这个概念。想象中的实体事物随着一般化事物的递增而递减。因此，数学符号是一种看得见的线索，帮助我们的意识完成从特例领悟通例的过程。

本书追溯数学中已确立的符号的起源和演化，始于计数，终于现代数学的基本运算符号。这主要是一部数学符号史；然而，它也探索了符号如何影响数学思考，以及符号如何唤起广泛又历久不衰的潜意识灵感。

本书包含三部分，区分数字的发展与代数的发展。这个艰难的写作决定是为了让可接受的符号定义能适合更大的记法范围，这个记法范围包括数字的记法和代数的记法。前两部分各有年表。第一部分和第二部分在某种程度上各自独立，但是读者应该知道，在早期发展阶段，数字与代数两者是沿着纠缠在一起的时间轴发展的。

---

[1] 关于所罗门圣殿中祭司行洁净礼用的水池，《圣经》中有一段引述（《列王记上》第7章第23节）：“他又铸一个铜海，样式是圆的，高五肘，径十肘，围三十肘。”这是对

“$\pi=3$”的诠释。但这里不是将π当成一种常数来引述的。

[2] 本书中所用的syncopate（简字化）这个动词，指借由省略一个词的中间字母来缩短那个词。这是特定的缩写形式，虽然大部分的缩写并非简字。19世纪中叶，德国数学家内塞尔曼（G. H. F. Nesselman）用三个阶段来描述代数记法发展的特性，他将这三个阶段依序称为文辞的（rhetorical）、简字的（syncopated）和符号的（symbolic）。

[3] 此处“类别”是指“ax^2+bx+c”这样的一般二次多项式，这是韦达的用词。——译者注

定义

> **符号**（sym•bol\′simbəl\（n-s）：由于关系、联想、习俗成规或偶然而非有意的类似，来代表或使之联想到其他事物的某种事物。

symbol是一个复杂的词。上述韦氏字典（Webster）所下的定义，并不完全与符号使用的集体经验吻合。为了更符合本书的内容，我们必须延拓前述定义，要让符号也是，或必须是某种具有文化意义且非随意定夺的事物，某种代表一个听起来或看起来并不相似的对象或概念的事物，以及对与它相似的事物不带先入之见的某种事物。

> **代数**（al•ge•bra\′al-jə-brə\（n-s）：数学的一个分支，在这个分支中，算术关系被一般化，并且在探究过程中使用字母符号来表示数字、变量或其他数学对象（如向量和矩阵），以及字母组合符号，特别是用以表示符合特定法则的方程式。

现今“代数”一词有更广泛的意义，扩及加法和乘法的通用规则，以及各种各样数学对象之间的结构关系。但由于本书主要讨论18世纪之前的代数学所使用的符号，因此韦氏字典的定义是恰当的。

关于插图

全书讨论了许多阐明符号史的原始信息，且在一些例子里以插图来呈现。尽管书中用以说明的一些原始手稿可以获得达到印刷质量的扫描文件，但碍于严格法律规定，以下页面中的文本必须重新打字：第65页、第100页、第137页、第149页、第150页、第151页。

第一部分 让人好奇的开端

主要的手稿与创始者

巴赫沙里手稿（*Bakhshâlî* manuscript，年代有争议：400—700年）

印度人所著。

显示印度人在公元700年之前，已建立一套位值系统。

塞维鲁主教（Bishop Severus Sebokht，约575—约666年）

叙利亚人，科学家、哲学作家。

已知在印度之外现存最早提及印度-阿拉伯数字的作者。

婆罗门笈多（Brahmagupta，598—668年）

印度人，数学家、天文学家。

著作《婆罗门修正历书》（*Brahmasphutasiddhanta*，628年）已知最早把零（一个小黑点）当作一个数字，而非仅是占位符号（placeholder）。

哈伦·拉希德（Harun al-Rashid，8世纪）

波斯人，哈里发。

建立巴格达的智慧宫（Bayt al-Hikma），一座图书馆兼翻译中心，藏有从许多其他文字译为阿拉伯文的数学和哲学手稿。

阿尔-花拉子密（al-Khwārizmī，约780—约850年）

波斯人，数学家、天文学家、地理学家。

智慧宫学者。830年撰写《还原与对消计算概要》（《代数》）［*Compendious Book on Calculation by Completion and Balancing*（*Algebra*）］。

马苏第（Mas’údì，Abu’l-Hasan ‘Ali，约896—956年）

美索不达米亚人，史学家、冒险家。

著作《黄金草原和珠玑宝藏》（*The Meadows of Gold and Mines of Gems*，957）是一部三十卷的文集，内容是波斯、印度、犹太、罗马和其他地区的历史，提供了10世纪关于九个印度-阿拉伯数字的可靠记录。

西尔维斯特二世（Gerbert d’Aurillac，Sylvester II，946—1003年）

法国人，教皇。

学习并教授数学，设计了一种称为吉尔伯特算盘（Gerbertian abacus）的计算板，有一套使用罗马数字的位值系统。

维希拉努斯抄本（Codex Vigilanus，约976年）

西班牙文本。

一份富有洞见的手稿，包含西方手稿中最早的阿拉伯数字。

伊本·以斯拉拉比（Rabbi Abraham ibn Ezra，1089—1164年）

西班牙人，天文学家、数学家。

著作《单位之书》［*Sefer ha-Ekhand*（*Book of the Unit*）］描述印度-阿拉伯数字符号，另一本著作《数字之书》［*Sefer-ha-Mispar*（*Book of the Number*）］描述位值系统和零。

切斯特的罗伯特（Robert of Chester，12世纪）

英格兰人，翻译者。

约1143年将阿尔-花拉子密的《代数》翻译为拉丁文，但19世纪才被发现。阿尔-花拉子密的《代数》是已知最早介绍印度-阿拉伯数字系统的文献之一。

希斯帕伦西斯（Johannes Hispalensis，John of Seville，12世纪）

西班牙人，翻译者。

著作《实用算术的计算书》［*Arithmeticae practicae in libro algorithms*（*Book of Algorithms on Practical Arithmetic*）］包含已知最早西方世界关于印度-阿拉伯位值记法的描述。

斐波那契（Leonardo Pisano Bigollo，Fibonacci，约1170—约1250年）

意大利人，数学家。

著作《计算书》（*Liber Abaci*， 1202）使用印度-阿拉伯系统，并以意大利商人的方言写成。

维尔迪厄（Alexander de Villa Dei，约1175—约1240年）

法国方济会修士和诗人。

著作《算法之歌》（*Carmen de Algorismo*）以拉丁韵文写成，说明包括零的印度数字的计算方法。

萨克罗博斯科（Johannes de Sacrobosco，1195—1256年）

英格兰人，天文学家、僧侣。

著作《算法》（*Algorismus*）是欧洲畅销教科书，讨论印度-阿拉伯数字和如何用它们做计算。

第1章 文明史上最重要的发明

没有人确切知道人类是从何时开始有意识地留下记号来与他人沟通的。可以确定的是那是在蒙昧不明的时期，成群的猛犸无拘无束地在欧洲大陆游荡，各种各样的生物循着食物和植被，从非洲平原向北分布。随着史上最大的气候变迁之一缓慢地结束，欧洲的冰原已消退了数世纪。多数人类族群仍分布在南亚。

5万年前至3万年前，当时的人类必须日复一日地思索如何获取生存所需。因此深层的本体论思想——诸如“我从哪里来？我究竟为什么存在？”——这些只能通过语言和隐喻来形成的思想，不可能存在。尽管没有充分发展的语言，他们必定有述说故事的本能欲望，那是一种将意识中存留的那些图像传达给他人的冲动。这些图像可能是关于暴风雨、黑暗、野兽的想象，甚至对梦境的困惑，但这些是推动语言进一步发展所需的养分。

随着语言的发展，对于生存体验的深思也在发展。20世纪杰出的民俗学者约瑟夫·坎贝尔告诉我们，人类总是“追求一种活着的体验，使得我们对纯物质层面上的生活体验产生共鸣，那些体验正是我们自身内心深处的存在与真实，以让我们实实在在感受到活着的狂喜”。

正如多数哺乳动物从过去到现在所经历的那样，人类无须发展出语言，借由本能与智慧的某种结合，便能够在艰险的环境中生存。他们可以在没有书写记录，没有记号、标志、符号或绘画的情况下，于一个口头交流的世界中度过严冬酷暑。猴子如此，驯鹿亦同。

是什么原因让那些新石器时代的洞穴画家着了魔似的忽视日常生活的危险，坐在那里铭刻、涂写或绘画？4万多年前，西班牙城堡洞窟附近的居民把手贴在穴壁上，对着手吹颜料，大费周章地摹印出手的图案。数万年来，人类在周遭留下有意义的记号，在树上凿挖，在硬泥地上留下脚印，在皮肤上刻画，甚至在岩石上着色。

一个简单的记号可以代表一种思想、说明一个计划或记录一桩历史事件。然而，关于人类的语言和书写最重要的事，莫过于说话者和书写者可以从有限多组记号和符号，创造出实际上无限多组读音、说

明、观念和想法。动物或许有自己的语言，但它们无法从有限数量的声音和姿态，创造出无限数量的沟通意符（signifier）[1]。

从岩石上的猛犸彩绘到字母表的诞生，文字的发展经历许多演变阶段。图画是图画文字的线索，图画文字又成为表意文字的线索，不断演变，直到成为早期隐喻性诗歌和现代文字的线索。“象形文字”是与它所要表示之物相似的图画。在亚洲，这类文字成为现代汉字的基础。在当今世界，一幅刀叉图画可以代表餐厅，画上斜线的刀叉图画可能是一种“表意文字”：它意指禁止饮食。象形文字描绘了物体，而表意文字则通过相似或模拟来表达意义。举例来说，早期的中国人为了以象形文字来表示“家”这个字，会结合代表“屋顶”的象形字符与代表“猪”的象形字符，创造出“家”这个字。至少有3万年的时间，故事是通过图画的方式来传诵，而随着时间的推移，故事变得越来越详细复杂。

几年前，一个朋友从泰国回来，送我一幅赫蒙族刺绣“故事织布”作为礼物，那是从难民营的一名织布工那里买来的。它描绘了越南战争时期日常生活的故事。从那幅刺绣中，人们可以“读懂”生命的循环。故事里有小孩诞生、在田地里工作、坠入爱河、婚礼和新生命——整个故事没有用到任何书写的文字。

象形文字让我们轻易了解一个简单世界里的简单故事。而当讲述的故事较复杂时，问题随之产生。想象一下用象形文字来“写”的《奥德赛》会是什么样子。谁能完全读懂它？用这种方式来理解太劳心费力，对于严肃诗歌所蕴含的复杂隐喻来说也很可能过于僵化。更好的做法是使用表示语音音素的字符，这样才能分辨一句话与另一句话——*a*表示“ah”，*b*表示“be”，依此类推。

拥有文字是一回事，思考文字本身的意思完全是另一回事。写下句子跟说出句子截然不同，它必然出现在社会显著进步之后，出现在第一个文明发展之后，出现在国王和君主之后，以及出现在富有冒险精神的部落开始漫游到熟悉的地点之外去探险和寻求贸易很久之后。

如果你在街上问一个人，文明史上最重要的发明是什么，你可能得到众所周知的答案：轮子。令人惊讶的是，轮子到了新石器时代晚期之前仍未出现，而且可能迟至青铜器时代早期。那大概是介于公元前6000年至公元前3500年之间。1976年在波兰布洛诺西（Bronocice）出土的陶罐上，可发现最早描绘带有轮子的马车的图画。这个陶罐的

年代可追溯至约公元前3500年至公元前3350年。但随着那个时期新农耕文化的发展，轮子应该是显而易见的发明——毕竟圆形的陶器和树干切面提供了极佳的线索，让人们了解滚动的圆盘状物体所拥有的巨大效用优势。可以确定的是，在真正的简易的轮子出现之前，人们已经学会了使用滚木。但轮子不仅是一个滚动的圆盘，它还涉及轮与轴结合这些相当复杂的概念。

那字母表呢？它肯定是轮子这个答案的竞争者。我将论述如果没有字母表或至少其他某种写下我们口头语言的聪明做法，大多数促进我们生活的重要发明实际上都将不复存在。诚然，那个路人或许认为，不借助滚木形式的轮子，奴隶无法建造出埃及宏伟的金字塔，而没有轮和轴，也无法打造出这个世界上高耸的石造建筑。轮子迟早会出现，但写下我们所发出的声音的某种形式，其重要性胜过一切。

现代的字母文字是对口语的一种粗略模仿。在关于字母表的任何证据出现之前，就有苏美尔人的楔形文字，其中苏美尔语里的每一个音节都是用楔形笔在黏土上刻下的独特图画。一开始，这些图画表示与所欲传达的文字发音相同的对象。这种书写方式与纯象形文字不同。幸运的是，口说的苏美尔语是一种由多音节词组成的语言，而且音节本身往往是具体对象的名称。这种文字由记号组成，每一个记号表示一个音节。举例来说，一个拿在手中的房子图画，可以意指“家庭”。

与苏美尔文字大约同时期，埃及周边的地中海地区使用象形图画文字，这种文字经历几个阶段的演变后，从图形字符悄悄演变为一种纯声音-符号系统，最终成为某种字母形式。到了腓尼基字母表传入时，大约公元前第一个千禧年之前的某个时间，世界上几乎每一个地区的众多文化都发展出运用图形符号表示文字的形式。这使文化交流成为可能，也为后世留下记录知识的工具。

和我们今日使用的表音文字不同，表音文字里每一个文字的符号代表口语的声音，图形文字是口语意义的指示物，而非声音的指示物。然而，到了公元前第一个千禧年中叶，图形文字被表音文字取代。

图画可以通过它们的声音来代表文字。在英文中，你可以借由并列一只眼睛（eye）、一只蜜蜂（bee）和一片叶子（leaf）的图画来写下“我相信”（I believe）。象形文字里的意义是通过上下文的语

境来表现，就像表音文字的情况一样。然而，表音文字至少有一个胜过图形文字的重要优点：它所能表达的思想与理念的组合要多得多。人们或许也认为，作家可以在自由得多的天地发挥，创造更丰富的隐喻。

书写的需求来自需要记录记忆而非记下故事，这点并不令人意外。最早的文件是记载账目、名字、食谱和旅程。当书写的技艺传播开来，书写的理由也越来越多。人们可以想象公共建筑物上的涂鸦、秘密笔记和传给其他人的神奇处方，书写帮助人们记忆，或为人们的墓碑写上墓志铭。这类记忆和墓志铭“唤起男男女女对于生命本身更深的觉察，引导我们经历从出生到死亡的种种试炼和创伤”。

刚开始，书写是仅限于受过初等教育者的活动，多半是神职人员或受过训练的特定阶级；然而，一旦建立了一些书写标准，它会给口语带来深远的影响。来自四面八方且拥有大致相似语言、受过教育的人，能够很快分享一种共同的书写语言，因而确立了语言惯例，并且在完全不同的地域和时代之间创造共同的经验纽带。

文明和城市的发端，与神庙的建造及神职阶层的崛起，具有惊人的一致性，而那些神职阶层是从寻常民众中招募的优秀人员。原始农业生活慢慢纳入神庙生活，神职人员建立了他们的帝国。这可能是农业文化发展所带来的结果，这种文化仰赖祭司解读历法，神庙也依据历法举行季节性仪式。因此，祭司——宗教活动或祭祀活动的神职人员——统治着最早的文明。他们的神庙是观测站、图书馆、医疗中心、博物馆和藏宝室。巴比伦人在公元前1200年制作了规模相当大的星辰表，而早在公元前3000年，埃及祭司已标出恒星和星座。星图涉及的复杂计算，还有土地测量和税务，需要的不只是点算田野上羊只所用的简单小数字，而是更大的数字。

原始人类需求简单。首先，计数限于非常小的数字。牧羊人不需要计数，就能知道羊群里少了一只羊。任何猿类都能做到这点——知道族群中少了一员。知道某个东西不见了是一种质性而非量化的集合观念。事实上，原始生活不需要任何真正的数感。没有人需要知道数字**是**什么。

但尽管如此，出于某种看来不可思议的原因，人类——甚至原始人类——总是有一种奇异的能力，可认知在他们的文字可表达的数值之外的数字。今天的学龄前儿童被教导要背诵数字，以了解与数量相

关的文字的意义。他们能轻易背出从1到10的数字。然而，背诵数字与了解那些数字的真正含义并不相同。一个三岁小孩不了解“一”“二”“三”“四”“五”这些字与一只手上的五根手指之间的一对一对应关系，或许仍能数到5。无论是发展于儿童或成人阶段，这种对应关系都是成熟运用智力的一个巨大进步。我们没有注意到这种飞跃发生的时刻。在那一刻，我们似乎不会有任何充满惊讶的顿悟体验。每只手各有五根手指这件事，似乎不会自然地让人联想到它与前十个数字的一对一对应关系。直到20世纪中叶，大洋洲几个原住民部落还没有表示数字的文字，却可以用在沙上做记号来计数。令人好奇的是，至少在20世纪之前，大洋洲、太平洋岛屿和美洲仍有一些原住民部落没有文字可以表示四以上的数字，代表一对一计数的现代数字概念尚未发展成熟。

无论在东方还是在西方，数学记录出现的年代都比文学早了一千多年。它甚至早于现存最古老的书写故事《吉尔伽美什史诗》（*The Epic of Gilgamesh*），这首苏美尔诗歌写作的时间比《伊里亚特》早了一千多年。我们没有直接的证据可以知道数字的书写最早在哪里或何时出现，正如我们没有直接的证据可知书写最早从哪里或何时开始发展。有些人认为数字书写最早期的概念源于中国人，最远可追溯到石器时代早期。这种说法似乎令人怀疑。但这样的说法与可追溯到公元前3400年的苏美尔人楔形文字的数字书写，或多或少有些巧合，因此显得合理。[2]

就像在法国南部和西班牙西北部洞穴中发现的艺术作品一样，数字的书写出于人类为了记录而做的努力。世界上现存最古老的书写记录之一（德国考古研究院博物馆，编号W19408，76+），看起来像是一道计算两块土地面积的练习，写于公元前第四个千禧年晚期。那是乌鲁克城（Uruk）废墟中被拿来再利用的建筑物石块里找到的一组泥板碎片，后来拼凑出这块泥板。它的碳年代测定（约公元前3350—公元前3200年）早于任何已知的书写证据，至少早于任何一种现代学者认为在语音上与口说语言相关的书写之前。

从欧洲到亚洲，在洞穴中发现的苏美尔人刻写在泥板上的数字记录，已经可以找到大如10000的数字。埃及象形文字曾用一个明确的符号来表示数字10000。到了公元前1600年，著名的莱因德纸草书中的代数问题提出一元一次方程式，除了用来表示数字的那些符号之外，方程式中没有其他任何符号。

---

[1] 原文signifier是“能指”的意思，此处译为“意符”。能指和所指是语言学上的一对概念。能指意为语言文字的声音、形象；所指则是语言的意义本身。按照语言学家或者哲学家的划分，人们试图通过语言表达出来的东西叫“能指”，而语言实际传达出来的东西叫“所指”。能指与所指之间的关系是自由选择的，对于使用它的语言社会来说，又是强制的。语言能指与所指的关系是非自然的，是可以改变的。能指同时既是意义又是形式，在形式方面它是空洞的，在意义方面它又是充实的。因为空洞，能指与所指的关系有偶然性，是约定俗成的。——编者注

[2] 根据美国考古学家、史学家詹姆斯·亨利·布雷斯特德（James Henry Breasted）的研究，历史上最早注有日期的事件（公元前4241年）是埃及历法。这个历法有十二个月，每个月有三十天，表面上比我们现行的历法好。不过这个埃及历法是不是真的那么古老，存在疑问。

第2章 古代人巧妙的计数办法

巴比伦人、苏美尔人、阿卡得人（Akkadian）——随你怎么称呼都行，我们早已听过他们的故事。早期西方的数学史几乎都从巴比伦人的数字概念讲起，一种所谓的六十进制（以60为基底），用以书写大数字、构成乘法表及研究天文学。但那些巴比伦人是谁，又为什么是他们最早发展出人类文明、文化、艺术和科学？

要回答这些问题，必须审视肥沃的弯月形地带，这个月牙形区域介于东地中海与波斯湾之间，并从土耳其东南部一直延伸到上埃及（Upper Egypt，尼罗河上游南方地区）。这恰恰是一个独特的地区，作为野生二粒小麦、野生一粒小麦和野生大麦的传播地，因此成为催生当地农业得天独厚的地区。特别是肥沃的弯月形地带有个地区靠近底格里斯河-幼发拉底河河谷。一般而言，Babylonian一词指称的对象远不只是巴比伦这个城市，基本上包含今日的伊拉克南部、科威特和伊朗西部部分地方的广大地区。这个地区位于两条大河之间，这两条大河在靠近现代巴格达的地方交汇，接着分流蜿蜒，直到流入波斯湾之前，在伊拉克南部的巴士拉（Al Basrah）汇流，该处恰位于科威特北方。如果看一下这些大河的地图，必然对它们的迂回曲折印象深刻。底格里斯河蜿蜒于巴格达南方，就像是一条无法下定决心要往西南还是东北游的水蛇。在一些地方——例如苏乌伊拉（Suwayrah）附近（参见图2-1）——乘船航行底格里斯河需两小时，走陆路仅需步行十分钟便能到达相同的地点。而在其他地方，走路半小时就可以到乘船得花六小时的地方。这意味着这条河两段非常长的河段之间的土地很容易灌溉，即使今日底格里斯河沿岸多数地方是未开发的农地。西方很少有陡弯急转的长河。一般来说，河流从高海拔地区往低处流。欧洲北部有陡然迂回曲折的河流——如上千公里的易北河——但北部的气候不太适合冬季作物。虽然底格里斯河-幼发拉底河河谷的地形不宜农耕，但大河的众多支流和水渠缓缓流过，非常有利于灌溉。小村庄沿着穿过巴格达南方的平缓的乡间河流形成，后来聚集成为西方最早的城市中心。最早的居民遍及这个区域时，巴格达南方冲积平原上有许多古老的支流和水渠，而今日它们已干涸消失了。

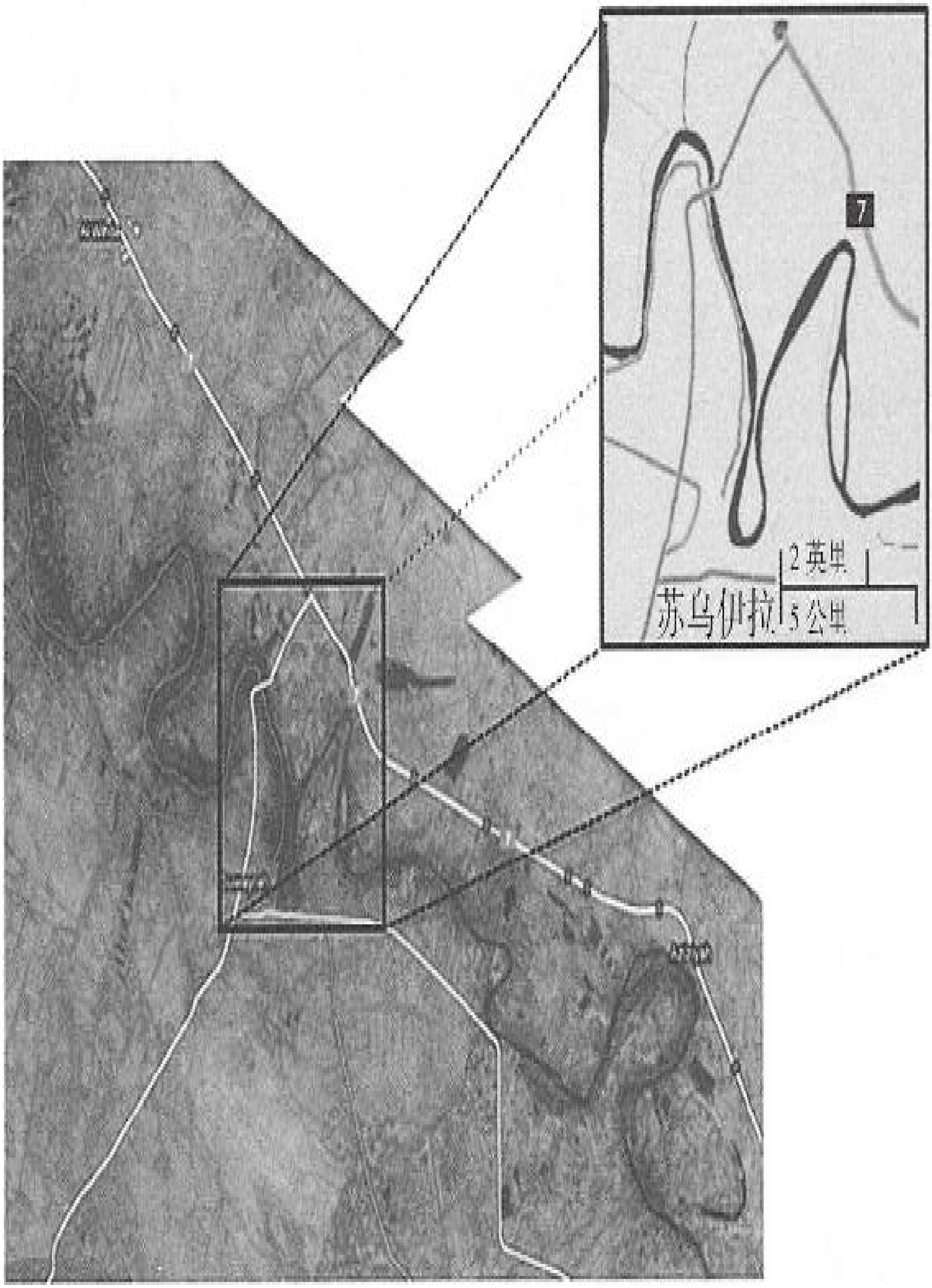

7
2 英里
5 公里
苏乌伊拉

图2-1 底格里斯河靠近苏乌伊拉的河段。资料来源：谷歌地图

然而，如果你活在汉谟拉比国王的时代——约3700年前——而且想定居下来种植作物养家糊口，你能找到的最佳地点是哪里？答案是南美索不达米亚。那里有平坦的沼泽地、大片肥沃的土壤和大量野生动物，非常适合种植大麦及畜养绵羊和山羊。那里是最早的城市文明落地生根的地方。

在城市化程度和有效率的农田灌溉方面，南美索不达米亚河流及水渠沿岸的帝国城市基什（Kish）、尼普尔（Nippur）、拉格什（Lagash）、乌鲁克、埃里都（Eridu）、舒鲁帕克（Shuruppak）、乌尔（Ur）附近，都大幅领先其他区域。巴比伦位于帝国的中心，这个帝国从南美索不达米亚延伸至幼发拉底河西北弯道。

除了受益于蜿蜒的河流和农耕季节长之外，应该有其他原因让这个地方显得如此特别。是土壤、贸易路线，还是祖先血统？根据《这些巴比伦人是谁？》（*Who Were the Babylonians?*）一书作者比尔·阿诺德的看法，原因既非土壤，也不是贸易路线。他论述道，埃及“与西亚其他地方大幅隔离，因为它受限于尼罗河谷所形成的狭长宜居之地”，也因此少有人入侵这块土地，这里的文化差异因而非常有限。而另一方面，美索不达米亚易受影响，因为这里几乎所有边界都一直受到不同民族所带来的丰富多样的文化影响。

巴比伦是古代世界的“熔炉”，不是因为欢迎外来者，而主要是因为它缺乏天然屏障，因此容易遭受外来入侵。南部地区广阔的平原和海湾水域都是容易进入的入口，东部和东北部的丘陵则是通往巴比伦城市中心的方便通道。半游牧民族的频繁入侵，让这里成为不断混杂交融、差异显著的多民族杂居之地。

南美索不达米亚大规模城市中心的不断发展，持续受到一个影响空前广泛的社会经济因素推动，史上第一次需要贸易和劳动方面的管理人员，以进行管治、组织和处理账务。这是之所以必须保存记录的原因，也是记账的开端。这些刻写在黏土上的账目是各种符号以及描述被记录的对象——土地、人物、家畜——的象形图案的组合。

20世纪初期，奥斯曼帝国正在瓦解，几乎无法管理次级古董的贸易活动（除了受贿与官僚掣肘之外），美国外交官、古董收藏家、小说家和巡游考古学家埃德加·詹姆斯·班克斯在公开市场上买了数百

块楔形文字泥板。后来他运送那些泥板到美国，卖给许多博物馆、图书馆和收藏家。他收藏的其中一块泥板让数学史家特别感兴趣。那块泥板是在森凯勒（Senkereh）发现的，这是古巴比伦城市拉尔萨（Larsa）和乌尔附近的考古遗址，位于伊拉克南部，为希伯来人始祖亚伯拉罕（Abraham）的出生地。

1922年，班克斯以10美元（根据美国消费者物价指数，约相当于今天的130美元）将这块泥板卖给纽约出版商乔治·亚瑟·普林顿[1]。要从历史的碎片来重现文化的片段总是困难重重，如何看待“普林顿322”（Plimpton 322）也就有了许多故事版本。1945年，数学史家奥托·诺伊格鲍尔和亚伯拉罕·萨克斯提出看法，认为那块泥板有毕氏三元数组表——也就是方程式$a^2+b^2=c^2$的一系列整数解。它之所以引人注目，在于它比西方所知的勾股定理起源的时间早了一千多年，而且说明了巴比伦人必定对勾股定理略有所知。然而，任教于剑桥大学的数学史家埃莉诺·罗伯森近来提出充分的理由，解释说那块泥板是教师的教具，用来讲解关于直角三角形的问题，根本不是勾股定理的雏形。

图2-2是一块巴比伦泥板上的笔墨图画，这块泥板是约3700年前于巴比伦重镇尼普尔古城制作的。那些记号并非小鸟的脚印，而是楔形笔刻出的痕迹。当把笔插入湿润的黏土板，就会留下一个形状或形状的印记。[2]接着会烘烤泥板。

检视左列，从上往下读。我们对这种古文一无所知，但可以猜想那一列代表数字1到12。左数第二列又如何？如果我们一开始的猜测是对的（怎么可能错呢？），那么我们可以知道那一列的第一个符号表示数字9。下一个符号会是多少？它必定是代表10的符号与代表8的符号并置而成。它会是18吗？同样的道理，第三个符号看起来代表27。嗯……第二列会是9的倍数吗？往下数到第六个看起来都是对的——也就是6×9＝54。到了第七个，似乎发生了奇怪的事。那个符号看起来像是4。但它是吗？如果是，第二列就不是9的倍数表。那它究竟是什么？

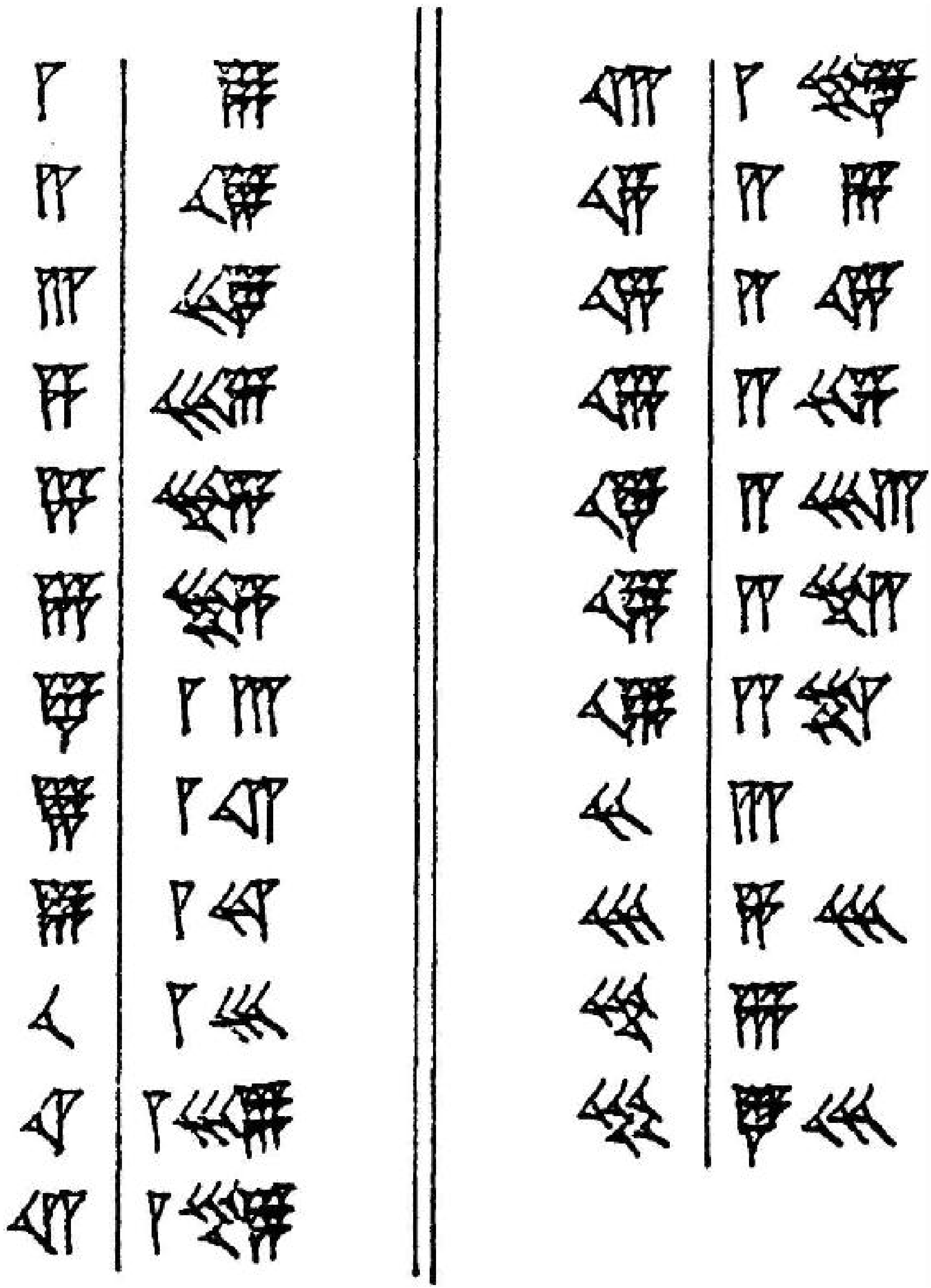


图2-2 尼普尔泥板。代表1，代表10。重绘自R.Creighton Buck，“Sherlock Holmes in Babylon，”*American Mathematical Society Monthly*， vol. 87， no. 5(1980)： 335-345. 经美国数学协会许可转载

我们注意到第七个符号左边的楔形记号与右边三个记号之间，有个小空格。如果认为第二列是9的倍数表，那么第七个符号应该是63。那个空格指的或许是，我们应该在加上那三个楔形记号所代表的数字之前，先乘上60。这样会得到正确的9的倍数。

以其他各项来验证这个假设，可以发现我们所做的推论行得通：

$$8\times9=1\times60+12=72$$

$$9\times9=1\times60+21=81$$

$$10\times9=1\times60+30=90$$

$$11\times9=1\times60+39=99$$

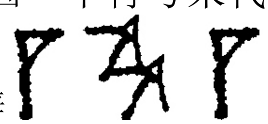
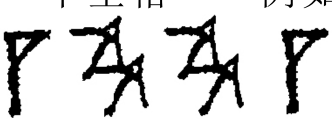
$$12\times9=1\times60+48=108$$

后面两列同样成立。所以我们在这里得到一个示例，说明很早以前就能巧妙运用记法，其中“空格”作为符号使用。

我们的数字符号（即我们现今的符号）与上述符号大不相同，而且精练得多。数字72表示7乘上10再加上2。我们需要的符号只有十个数字（0，1，2，3，4，5，6，7，8，9），这样便可以表示出我们想表达的任何数字。尽管以门外汉的眼光来看，巴比伦系统似乎需要五十九个不同的记号，但其实只需要两个符号。为了表示小于60的数字，代表较小数字的符号有条理地放置在一起，几乎是一个接一个。举例来说，数字39会写成：

我们把六十一写成61，而且知道它指的不是3601。但巴比伦人如何区分61与3601？数字61表示成，数字3601则表示成。两者唯一的区别是分隔楔形记号的空格数量（中间有一个空格，而中间有两个空格）。然而，因为看不出空格的边界，很难知道记号之间有多少空格（特别是当那些记号是手写时更是如此）。这是件麻烦事。一个空格就是一个空位，而两个空格看起来可能还是像一个空位。

你或许认为，根据上下文中数字相对值的大小可以区别不同的数字，就像对上下文的理解有助于厘清语言的含混情境一样。举例来说，要区别 goats与 goat很容易：第一个记号暗示了复数，因而必定表示60只山羊，而第二个记号暗示是单数，必然表示1只山羊。根据上下文可以区分（3601）与（216001）吗？说不定可以。

必须设计出一种工具，使这套系统得以运作。从21世纪的后见之明来看，我们可以清楚了解那个工具是什么。随便画一个符号来代替一个空格——例如，那么读者便可轻易分辨与。那为什么没有这样做呢？

曾经有人做过。然而，和罗马一样，但与现代的北京不同，它不是一天造成的。必须有人想出一桩妙计。事实上，这花了一千多年的时间才实现。公元前700年至公元前300年间，有人想到了利用一个看起来非常像的符号来表示一个空格。这是占位符号的发明，也就

是巴比伦人的零，尽管它并不符合现代的零的意义。从此之后，可以在不是仅依赖上下文的情况下区分与。

虽然我们必定觉得这套系统看来奇怪，但它是一套杰出的系统。只要能辨别空格，巴比伦数学家就可以只用两个符号和那个区别出空格的奇特记号，来写出任何数值的数字。[3]

远在巴比伦的抄写员于尼普尔炽热的阳光下把芦苇茎压入黏土中之前，埃及人已在石头、金属和木制纪念物上雕刻象形文字。在那个时期，数字仍只是对象的简单图画，那时十的每个乘幂都用不同的符号来表示。数字1以一根垂直的棍棒来表示；数字10以一根弯曲成半圆形的棍棒（轭）来代表；数字100是一个蜗牛图形；数字1000是一朵莲花；数字10000是一根指示的手指；数字100000是某种看起来像小鸟或可能是鱼的东西；数字1000000是一个举起双手的人，仿佛因数字的巨大而手足无措（参见图2-3）。

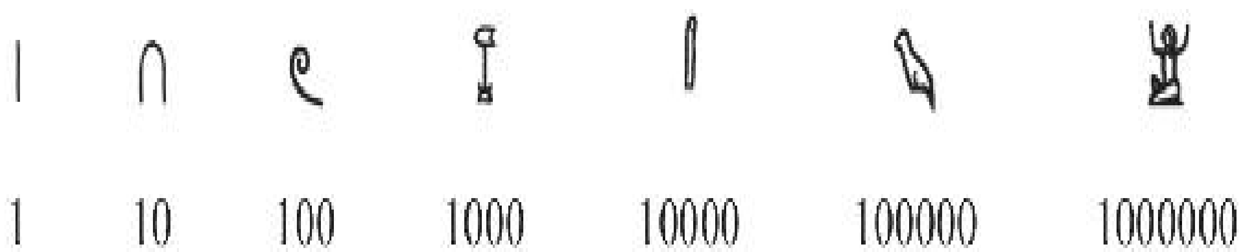


图2-3 早期埃及的数字书写。引自Florian Cajori，*A History of Mathematical Notations*（New York：Dover，1993），12.

有些埃及学家推测，蜗牛图形是一根卷绕的绳子，那只小鸟其实是青蛙，而那个看起来足以代表大如一百万这个数字的人其实是神。此外，因为埃及人不需要任何更大的数字，例如一千万，那个神也代表任何非常大的数量。这种数字书写是依循加法原则，永远将表示较大数字的图案放在表示较小数字的图案左边。举例来说，数字3601（从右至左）会写成一根直棒、六只蜗牛和三朵莲花的图画。这个体系运作得很好，不需要占位符号，这是明显优于巴比伦系统之处。数字3601与36001截然不同。要写36001，不需要绳子，只需要一根直棒、六朵莲花和三根手指。61与3601不会混淆。再者，36与3600也不会混淆，因为后者会用六根绳子加上三朵莲花来表示。

早期埃及的数字书写是一种加法系统。要写1005，人们只需要在五根直棒旁加上一朵莲花；然而，不能超过四根直棒，所以抄写员会把那些直棒分成两堆。在公元前第二个千禧年之后，乘法系统诞生了。要写数字2000000，抄写员会在草纸上画一个头顶两根直棒的人。但仍有许许多多谜题是埃及学家无法解答的。举例来说，在象形文字中单位分数$\frac{1}{2}$写成图案

，$\frac{2}{3}$写成

。

希伯来人有不同的体系。他们的字母表有二十二个字母，每一个字母都表示一个数字（参见表2-1）。另外还有五个字母只用在词尾。这些字母是ץ，ך，ם，ן，ף，分别代表500，600，700，800和900。

表2-1 希伯来字母表

א	（Aleph）	1	ל	（Lamed）	30
ב	（Bet）	2	מ	（Mem）	40
ג	（Ghimel）	3	נ	（Nun）	50
ד	（Dalet）	4	ס	（Samekh）	60
ה	（He）	5	ע	（Ayen）	70
ו	（Vav）	6	פ	（Pe）	80
ז	（Zayin）	7	צ	（Tsadi）	90
ח	（Het）	8	ק	（Qof）	100
ט	（Tet）	9	ר	（Resh）	200
י	（Yod）	10	ש	（Shin）	300
כ	（Kaf）	20	ת	（Tav）	400

要表示千位数，人们会从第一个字母开始，在那个字母上方放两个小圆点。因此，א̈表示1000，ב̈表示2000，依此类推。现在这里

有个诀窍。希伯来文是从右读到左，大于一千的数字会有两种写法。正如这几页讨论的所有希伯来数字体系，每一个文化的数字体系都历经数世纪的尝试和变革。到了8世纪，双字母符号ה̈א表示5001。字母א原来表示数字1，但当它出现在另一个字母右边时——例如ה——是表示1000。这不会产生混淆，因为即使这些字母从右读到左，它们的数目等价结果可以理解为数值递减。因此，ה放在א的左边时，表示的是1005。

这个体系运作得很好。数字9686可以写成טוחו。请注意，字母ו出现了两次，而它被视为两个不同的数。它原本表示数字6。从右读到左，第一个ו的值必定介于ט与ח的值之间，因而必然表示一个介于80至9000的数字，因此，ו一定是600。剩余那个位置上的ו一定代表它所能表示的最小数字，就是数字6。

一般而言，符号的本质是将彼此无关联的事物的意义连接起来，以创造一种思维状态。在希伯来文中，数字15原本可以从右到左写为י（表示10的符号）加上ה（表示5的符号）。然而，以这样的方式写15，也会写下神名的前两个字母[4]。所以数字15当时（且至今仍是）写成6＋9（טו），而不用上述写法。

希腊人借用了希伯来系统来表示数字。他们同样以希腊字母表的字母来分别代表各个数字——这种体系在表示大数字的时候极为不便。

α β γ δ …

1 2 3 4 …

为什么他们不借鉴巴比伦系统，利用占位符号和相对简单的方式来写大数字？巴比伦人已有位值记法（positional notation）的正确概念，也就是用相同的数字来表示60的各种不同乘幂这个巧妙的想法。那些数学资源丰富的希腊人怎么会没注意到如此具有启发性的想法？从他们发展出来的一切——他们对逻辑思考、证据和证明的组

织，他们对几何和无理数的理解，他们借由几何来解决数论问题的能力——为什么他们没有发现更好的数字处理方式，让算术变得更容易？一个答案可能是，某种形式的计算器大量应用于计算。

或许因为他们感兴趣的是领悟数学本身的广大领域。计算不是他们真正想追求的，尽管无疑有很多数学家也从事非演绎的数学。他们发展的是一种严格的演绎科学、证明、求解、普遍、完美，以及理解欧氏空间和充满那个空间的对象之关系，而所有这些都是利用一种相当笨拙的数系，以及在几乎完全不需要数系的精练层次中完成！

公元前8世纪，希腊人采用了腓尼基字母表，并利用它的“首字母”，从表示数字的书写文字的开头字母衍生出符号（参见表2-2）。

表2-2 希腊首字母系统

1	I	Ἰώτα
5	Π	Πέντε
10	Δ	Δέκα
100	Η	Ἡκατόν
1000	Χ	Ξίλιοι/χιλιάς
10000	Μ	Μύριον

备注：要写5乘上一个数，把该数放在雨伞符号Π下。举例来说，𐅅表示500。

公元前5世纪，希腊人开始与希伯来人、叙利亚人及腓尼基人开展更普遍的贸易活动，而他们各有自己的字母表，于是慢慢形成了与其他系统竞争的局面。当这种情况发生时，那些擅长字母表的腓尼基人借鉴了埃及象形文字标志，赋予它们独有的读音，然后将那些读音用字母来表示，但奇怪的是他们没有用自己的字母表来表示数字，而是用一种垂直线条的系统。

公元前4世纪之后，希腊人的字母表序列数系脱颖而出，取代旧有的首字母系统。就像希伯来系统一样，希腊字母表系统成为标准规范。

要用这个系统来表示大数字是非常笨拙的。即使是阿基米德，当他撰写具有独创性的著作《数沙者》（*The Sand Reckoner*）时，关于

填满宇宙所需的沙粒数量的估算，凭借的是文字而非记法，才得以描述这么大的数。利用我们的记法来表示，他的解答是该数接近10^{51}，远少于更正确的答案：约10^{90}[5]。

但这个字母表序列数系为何受到青睐？为什么希腊人舍弃首字母系统，转向字母表系统？只因为它能较简短地表示出数字吗？数字1884的首字母记法是Χ𐅅ΗΗΗ𐅄ΔΔΔΙΙΙΙ；而用字母表记法，1884写成αωπδ。记住二十七个符号所代表的数字，不是比只记六个符号难吗？是的，但随着时间过去，儿童也能记住这些数字，就像记得字母的顺序一样。

但情况并非如此。两种系统似乎存在概念上的差异。20世纪早期的数学史家弗洛里安·卡裘利在他成就极高的数学记法史研究中，考虑了下面两个算术等式（请牢记，加号和等号在16世纪之前尚未出现）：

ΗΗΗΗ + ΗΗ = 𐅅Η

ΔΔΔΔ + ΔΔ = 𐅄Δ

上述等式用字母表示为：

υ + ς = χ

μ + κ = ξ

这里出现了选择。两种表示方法都是人类创造的，它们竞相争得必须使用它们的算术家和抄写员的青睐。两者都很难处理，只有一个胜出。为什么熟知巴比伦系统的希腊人，没有想出一种更聪明地运用占位符号的系统？为什么留给旁遮普东边的印度人来提出所有数系中最巧妙的一种？

我们可以发现，图2-4中隐约暗示了一种位值系统。前十个希腊字母代表前十个数。要表示从11到19的数，人们会写成ια，ιβ，ιγ，ιδ，ιε，ιϛ，ιζ，ιη，ιϑ，代表10＋1，10＋2等，依此类推。然后，要表示20到29，可以写成κ，κα，κβ，κγ，κδ，κε，κϛ，κζ，κη，κϑ，代表20，20＋1，20＋2等，依此类推。虽然里面有一些发明出来的符号，例如表示90和900的怪异符号，但创造出这套系统的人们应该已了解一个符号的放置（placing of a symbol）可标示出其值。若要在希腊人的系统里写数字23，因为那个系统中位置并未标示出值，迫使人们必须引入一个新的符号（κ）来表示20。而要在一个位置标示出值的系统里写数字23，需要的只有β，这个符号表示2，一个已定义明确的符号。当β出现在第二个位置时，它代表20，而不是2。数字23可以写成βγ。我在这里提出这点，是要说明符号系统中的记法选择，可能成为未来发展的阻碍。就像其他任何字母表数系一样，希腊体系处理小数字没问题，处理大数字却很笨拙。

α	β	γ	δ	ε	ϛ	ζ	η	θ	ι	κ	λ	μ	ν	ξ	ο	π	ϙ
1	2	3	4	5	6	7	8	9	10	20	30	40	50	60	70	80	90

ρ	σ	τ	υ	φ	χ	ψ	ω	ϡ	͵α	͵β	͵γ,
100	200	300	400	500	600	700	800	900	1000	2000	3000

etc.

M	$\overset{\beta}{M}$	$\overset{\gamma}{M}$,	etc.
10000	20000	30000	

图2-4 希腊字母表序列系统。表示6的字母是ϛ，一个草写体的*digamma*，这是一个古代字母，公元前7世纪之前从希腊字母表中消失。它看起来像用在词尾的sigma，但它的读音非常不同。当我们在第二部分谈到作为非数字符号的另一个sigma时，务必牢记这点。注意6是以ϛ来表示，即使按真正的字母顺序应该以ζ表示6

这些古代的字母表不仅是成堆的各具特性的具体语言元素，也是多重意义日趋成熟的基石。公元前5世纪，希腊人相信世界上所有事情都可以联系到整数。数字2（字母β）意指看法，3（γ）是和谐，4（δ）是正义。奇数为男性，偶数为女性。数字5（ε）象征婚姻，可能因为它是第一个偶数与第一个奇数之和。而数字10（）是神圣的，因为它是前四个维度之和（点、线、三角、四面体），即1＋2＋3＋4＝10。于是我们开始了解到，这些古代数系激发了各种各样的隐喻性的思维状态。

罗马系统与希腊首字母系统密切相关，它们使用加法原则来表示大数字，并聪明地运用减法规则：当一个较小的数放在一个较大数的左边时，它指的是由大数字减去小数字（参见表2-3）。

因此，83可以写成XXCIII，而不是较长的表示法LXXXIII。理想的数字语法要求以可能的最短长度来表示数字。这个规则有很多变化。举例来说，到了4世纪，我们发现一些地方的作者会在数字的上方画上一条水平线，表示该数的千倍，所以$\overline{\mathrm{X}}$表示10000，而不是10。在一个数字的左边和右边加上垂直线，上方加上一条横线，表示的是该数的十万倍，所以$|\overline{\mathrm{X}}|$表示1000000。就像希腊体系一样，用这种方式来表示大数字极为不便。我们现在仍然用罗马数字来表示日期，可是我很想知道个中原因。

表2-3 晚期罗马数字符号

1	I
5	V
10	X
50	L
100	C
1000	M

虽然阿兹特克数字与亚洲、非洲、欧洲的数字没有直接的历史渊源，但它们有相似之处。阿兹特克数字先以小圆点来代表一直到9的单位；9之后的数字变成图画形式。一根完整的羽毛表示400，所以四分之一根羽毛是100，半根羽毛是200，四分之三根羽毛是300。表示8000的符号是一个小袋子，大概装了20乘400，虽然那个袋子本身没有清楚地表示出那个积（参见图2-5和图2-6）。

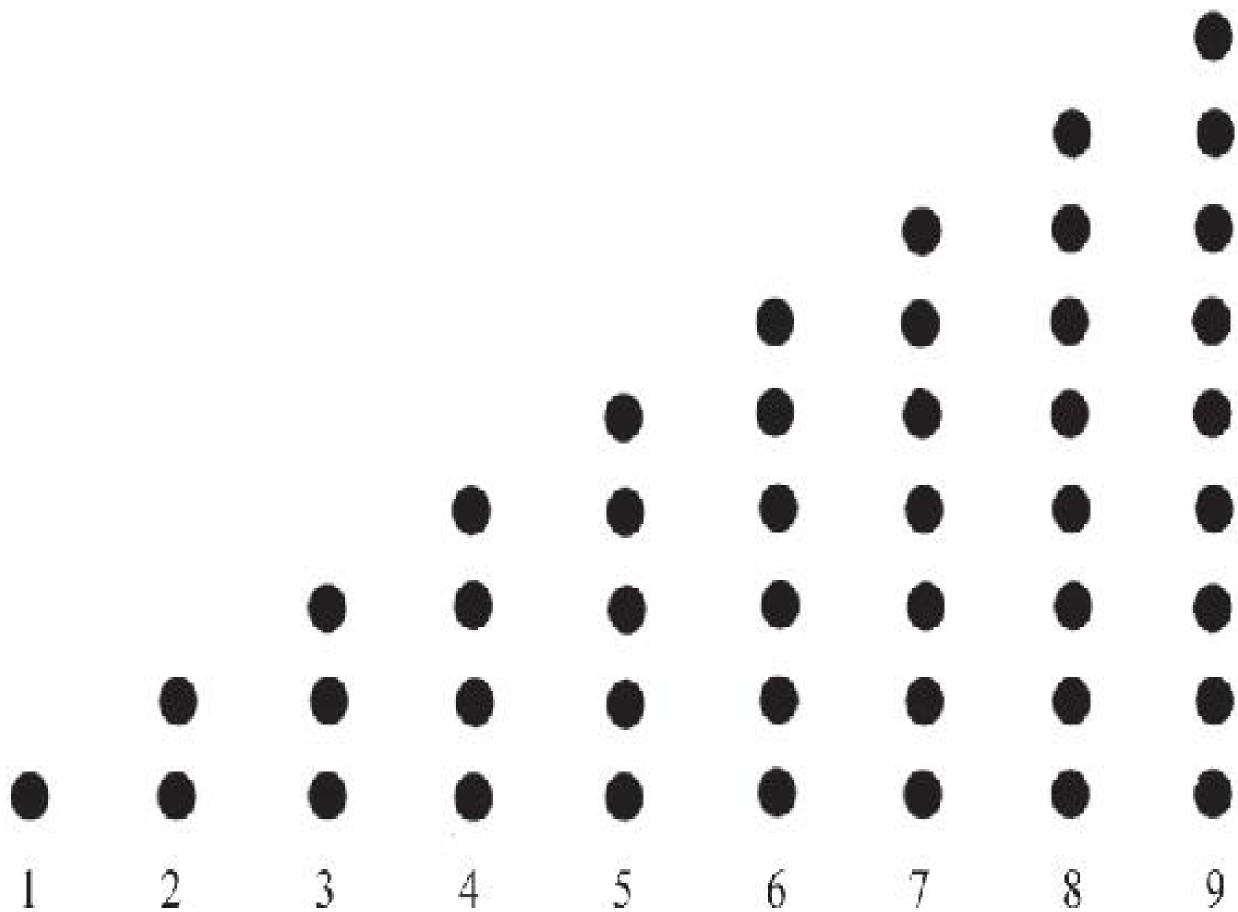


图2-5 阿兹特克数字（较小的数字）

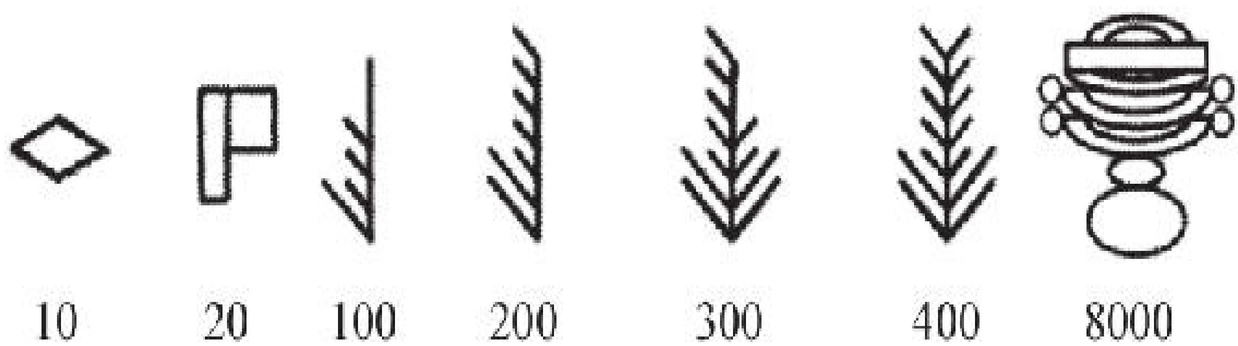


图2-6 阿兹特克数字（较大的数字）

如同其他大陆所发明的系统一样，这个系统是依循加法原则。但与巴比伦的单一基底系统不同，阿兹特克系统有三个基底：20，400和8000。例如要写26504，阿兹特克人会写成如图2-7那样。

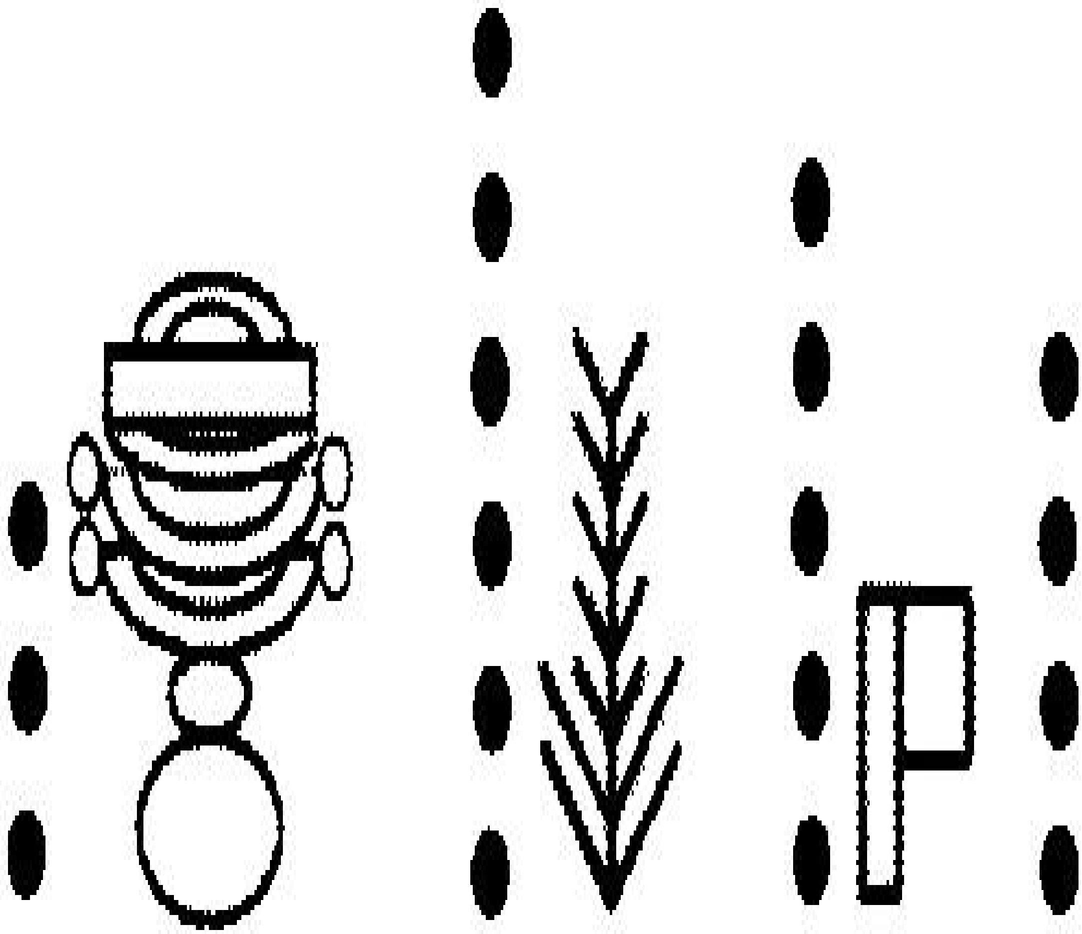


图2-7 阿兹特克的26504写法

玛雅人（年代不详，但可能在古典期，250—900年）则采用一种类似二十进制的系统（以20为基底）。这种玛雅算术系统是前哥伦布期（pre-Columbian）的发明，而且新旧两大陆超过五万年不曾有人类接触，但是它们对加法与进位符的观念相似。这种系统类似巴比伦系统，它在点、横线和列组成的系统中，运用了零的占位功能。一个点表示一个单位，一条横线表示五个单位。举例来说，3212119会写成如图2-8所示。

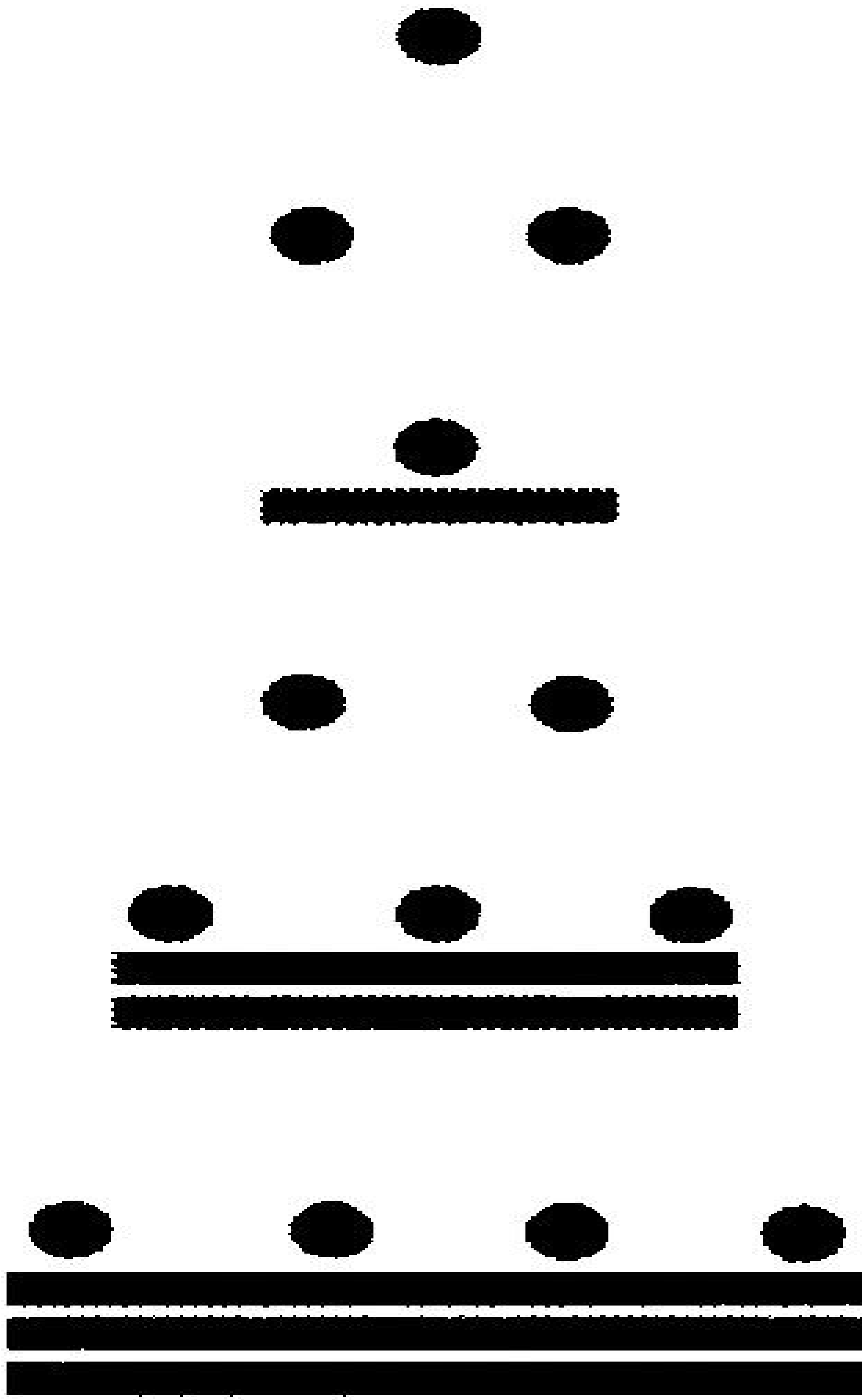


图2-8 玛雅数字

从上至下表示：

1乘以18×20×20×20×20

2乘以18×20×20×20

6乘以18×20×20

2乘以18×20

13乘以20

单位层（19）

由下至上的和为：

19＋260＋720＋43200＋288000＋2880000＝3212199

这个系统让算术变得容易。要将两数相加，把它们写成两列，将它们各自的行相加，然后把所得的列加起来，采用进位的方式，就像我们现代加法系统的做法。举例来说，要把55和151相加，玛雅人会写成：

151　　55

接着把每一行的数字加起来，得到：

206

最下方数字的四条横线会“进位”，成为下一层的一个点：

206

数字206很容易写，较小的数字如20反而麻烦。当单位层没有任何东西时，玛雅人不能直接在第二层放上一个点。所以他们巧妙设计了一个符号来代表该层为空，一个看起来有点像的零。于是数字20可以写成。

---

[1] 它是普林顿的私人收藏品，直到他于1936年去世为止。后来它被遗赠给哥伦比亚大学，现存于楔形泥板藏室。

[2] 早期的楔形标记是以芦苇刻成的，留下圆形或半圆形的印记。

[3] 在某种程度上，只使用了两个符号和。其他五十八个符号可视为这两个基本符号的组合；不过，我们往往会将每一个组合视为不同的符号。

[4] 希伯来《圣经》中以四个希伯来子音 יהוה 来尊称神名，即“四字神名”：YHWH（中译名为“雅威”）。文中说“神名的前两个字母”是指从右往左看。——编者注

[5] 这个数经常与阿里斯塔克斯（Aristarchus）后来利用阿基米德的方法所求得的估计值10^{63}混为一谈。

第3章 不得不佩服的中国人

孙子曰：夫算者……稽群伦之聚散，考二气之降升……

——《孙子算经》序[1]

地形因素加上马不停蹄地频繁往来，形成连接中国到印度与印度到波斯的东西方路线。这条路线不是筑路工人开凿而成。丝路不是一条特定的道路，而是一连串海陆路线，纵横交错欧亚，穿越六千多公里荒芜崎岖的地势，并连接其他主要是印度商人、掮客和探险家行经的路线。这条路线在公元前2世纪左右形成，连接了波斯札格罗斯山脉（Zagros Mountains）的波斯御道，邮驿人员在这条路线上传递邮件，人们也可以找到精力充沛的马来一路走到地中海。缎、丝、大麻、香水、香料、珠宝、玻璃和药品运往西方；金、银、地毯和酒运往东方。就像所有主要的商业贸易路线一样，丝路和波斯御道也是文化、宗教和哲学思想的传播路径，以及各种轻疾重症的病菌的传播路线。

哲学、科学和数学的知识，同样经由这些跨国要道传播。商业交易大多通过以物易物的方式进行，但公平的以物易物至少需要约略估算价值，了解重量与度量的换算：丝的正方形面积、金的重量或钱币的币值。一个与波斯掮客及中国掮客交易的印度人必须了解商业数学，并且能够传达及理解某种数值信息，而这可能是通过对西方数字与东方数字的描述之间的转换来完成的。

在几百年间，中国数学史已大半佚失或被毁，主要是暴君下令焚书之故，这使得西方人容易相信西方主宰了数学的起源。最古老的中国数学记录，包括最早的书写数字，可追溯至商代（公元前1600—公元前1046年）。1899年，考古学家在曾为商代首都的中国河南安阳小屯村遗址，挖掘出数以千计的骨头和龟甲。[2]此后，收集研究了成千上万新发现的甲骨。上面的数字符号记录了在战役中虏获或杀死的敌人人数、猎获的鸟兽数量、献祭的动物数量，以及许多其他功绩。

到了西汉（公元前206—公元9年）初，中国发展出了十进制计数汉字，看起来非常像今日所用的中文数字（参见图3-1）。

1	2	3	4	5	6	7	8	9	10
一	二	三	四	五	六	七	八	九	十
10^2	10^3	10^4							
百	千	萬							

图3-1

举例来说，数字26999写成：

二萬六千九百九十九

我们必须称赞这个系统的精巧程度。它从左到右念成：2万，6千，9百，9十，9。毫无疑问，这是一种十进制系统。为什么它如此精巧？它完全不需要零，至少不需要以零作为占位符号。要写20009，只需要写：

二萬九

关于这点，我们真的必须佩服中国人。没有十位、百位或千位时，不要用它们就好。不需要零！

我想数学史家如果发现是哪些聪明的中国数学家提出如此卓越简洁的想法，肯定会欣喜不已，这不仅是极好的商业书写数字系统，对于土地测量和天文学来说也是好用的概念。可惜的是，由于战争、焚书和手稿损毁，连主要是哪些人提出这个概念都几乎一无所知。

这种巧妙的数字系统让中国人得以“命名”大数字，但要做实用算术还需要其他某种东西。于是，中国人又发明了一种了不起的系统：算筹。

远在第一个千禧年之前，中国人就普遍运用兽骨或竹子做成的算筹来代表数字一到九（参见图3-2），这个系统是一个以十为基底的位值数系。那是一种非常像我们今日所用的十进制的纵筹和横筹组合，只是其中仍然没有零的占位符号观念。

图3-2 中国算筹[3]

命名数字和手指计数在表示小数量时可能没问题，但进行加、乘、除需要做某种移动和移除：书写、划掉和改写。在公元前1世纪没

有便宜纸张的时代，一连串计算过程中可以快速移动和移除的算筹是最有效率的方式。就像印度-阿拉伯系统一样，中国的书写数字与算筹数字都是位值制，而且使用方便，不只用来表示数字，也有助于计算和数学概念的理解。

算筹使用方式的知识，在汉朝（公元前202—公元220年）的《九章算术》手稿中已出现，这部收有两百四十六道问题的巨著写在传统的竹简上。史学家认为这是中国最早完整收录此前所有已知数学知识的文本，一本货真价实的中国版欧几里得《几何原本》，其终章——即使不是欧几里得公理逻辑系统意义上的证明——提出一个对于勾股定理的严谨且可理解的说明。中国人研究数学的方式是，借由例子、模拟、不亚于欧几里得有效性的证明（或至少是说明）来展开。

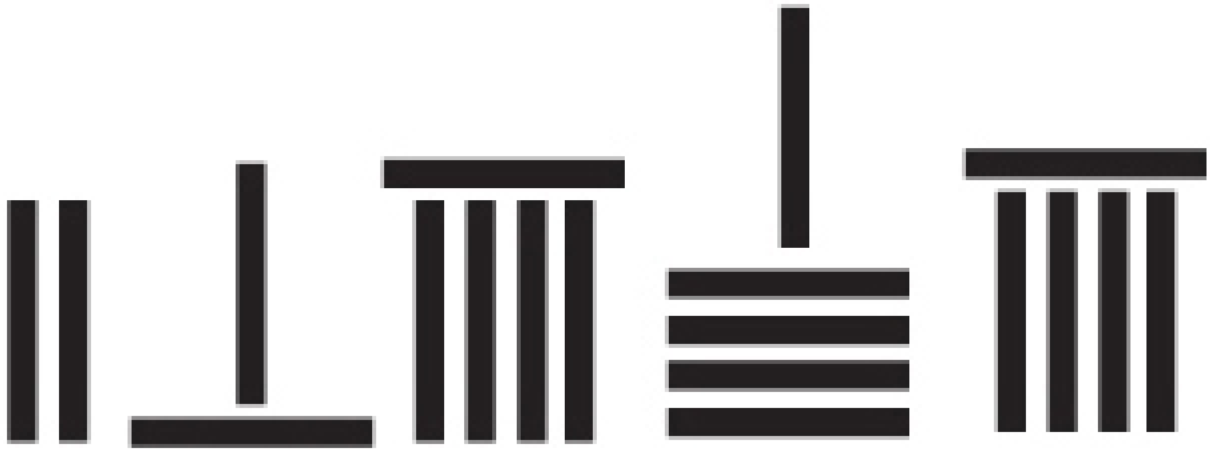
不幸的是，它就像公元前第一个千禧年的几乎所有的书一样，没有原书的完整复本留存，可能是毁于秦始皇以弃旧迎新的荒谬托词下令焚书，虽然更可能的原因是，就像所有缺乏安全感的暴君一样，秦始皇想删除任何会把他的统治拿来与过去的皇帝做比较的证据。然而，幸运的是，我们有公元263年版的《九章算术》，这是张苍和耿寿昌在公元前1世纪编纂的。刘徽为该书作注并增补。他在序中写道：

徽幼习九章，长再详览。观阴阳之割裂，总算术之根源。

两位杰出的中国数学名家——新加坡国立大学的蓝丽蓉和澳大利亚伊迪斯科文大学的洪天赐，在两人合著的《雪泥鸿爪溯数源》（*Fleeting Footsteps*）中告诉我们，印度-阿拉伯数系可能源自中国算筹的推测不是毫无根据。

算筹可以想成摆放在一个平坦表面，也就是一块计算板上的牙签。许多早期的中国数学文献，让我们得到关于中国算术的一些启发——特别是公元四五世纪的《孙子算经》，书中以纵筹和横筹来表示数字。红筹为正，黑筹为负；把算筹放进（或移出）计算方块，以进行加减乘除的运算。

算筹是一种位值系统，举例来说，数字26999写成：



这与我们的印度-阿拉伯十进制系统极为相似。问题出现在想表示2600999这样的数字时。这个系统还没有零的占位符号观念。原始的印度-阿拉伯十进制系统没有代表零的符号，但的确有一个表示空位的字（印度文是*sunya*，阿拉伯文是*.sifr*）；中国算筹也是利用一个空格来作为占位符号，中文以“空”这个字代表空位。就像巴比伦系统的做法，空位有重要意义，但用手写表示时，空白处往往会产生歧义。图3-3的中国算筹数字指的是2600999还是260999？

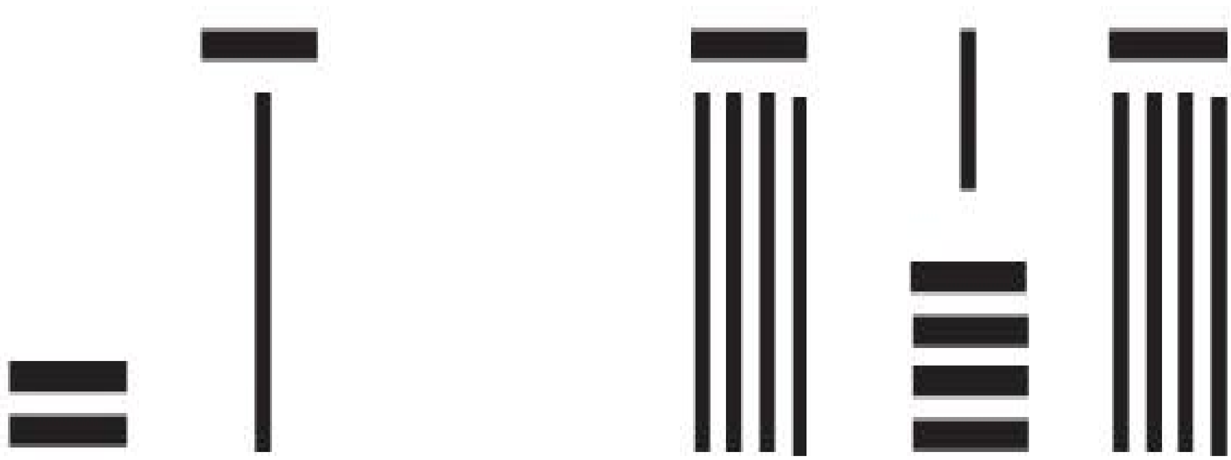


图3-3 2600999还是260999？

关于这点，中国人同样富有创造力。请注意，万位的6与百位的9之间有一个空位（一种零）；然而，可以确定这个数字是260999而非2600999，因为算筹的摆放方向随不同位数交错。表示6的算筹与表示9的算筹方向相同；中间跳过的部分意味着6与9之间有一个零。要写26000会有一点小麻烦，但简单的解决之道是写成一千，解释为“26千”。

还记得古巴比伦系统如何用一个空位来作为表示零的占位符号，以及用两个空位来代表两个零吗？中国的系统巧妙利用摆放方向交错的方式，避开了这种情况。因此，若两个零被夹在6与9之间，6与9的摆放方向就会显示出来。

多么聪明呀！这是正反阴阳的观念。这种有效的小伎俩也有助于区别两个零与一个零。图3-3所示的数字是260999。数字2600999以不同的写法表示：

看起来6与9之间似乎有两个空位，这让此数为2600999，但我们是检视算筹的交错摆放方向来确定这一点的。

如果6与9之间只有一个空位，2和6的摆放方向应该与图3-3所示的相同。这个系统在区分26990与2699时特别有用。前者表示为：

而后者为：

另一部重要的数学著作是《筭数书》。1983年，考古学家在中国中部发现的一座古墓中找到包括该书在内的文物，这批文物自公元前1世纪尘封至今[4]。当时发现了近两百片竹简，连接起那些竹简后形成《筭数书》，里面有同样精巧的算筹，但多了某种东西：一种用于算术计算的矩阵系统[5]。图3-4解释了如何以9除6538。

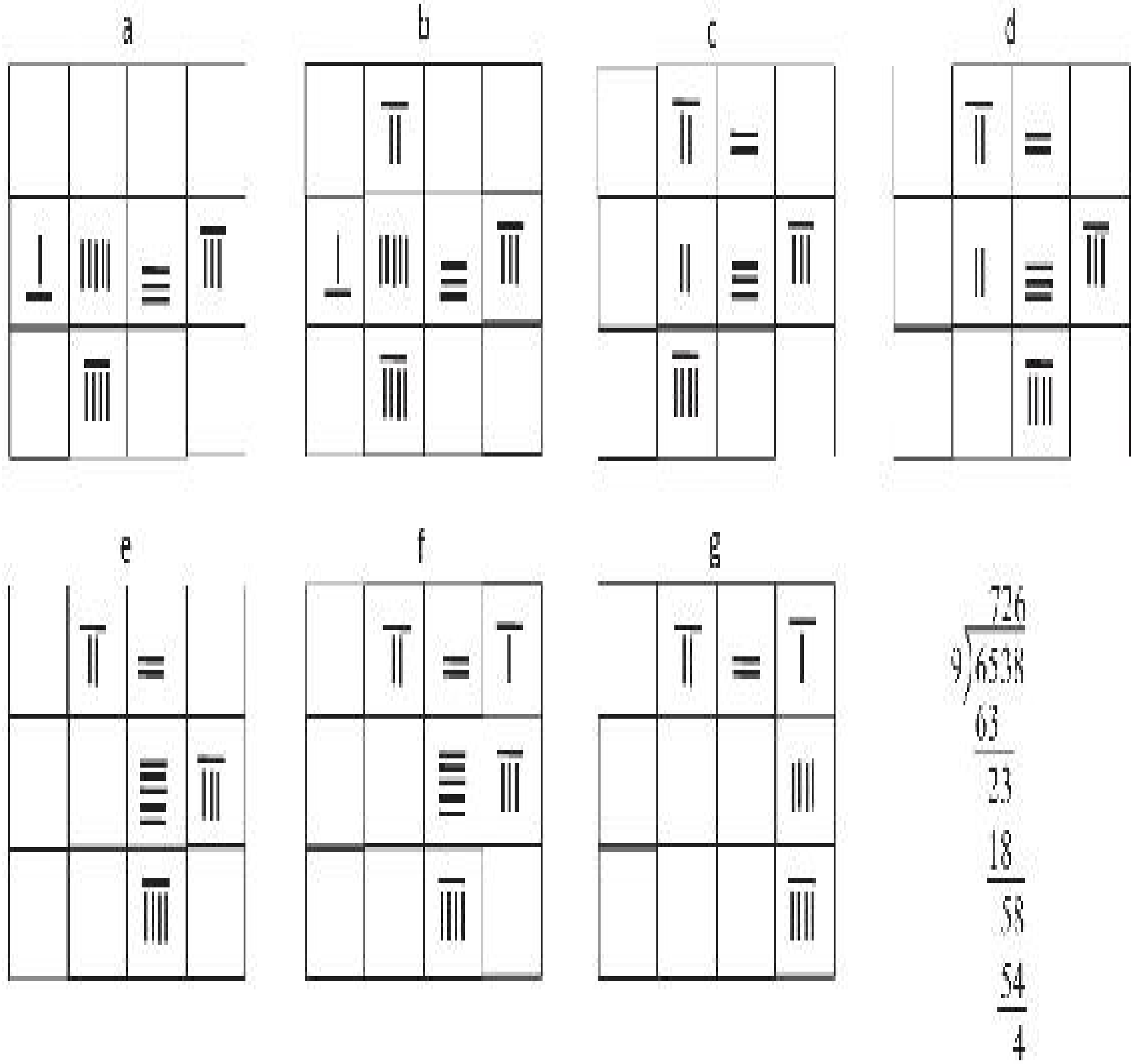


图3-4 以9除6538。资料来源：Philip D. Straffin Jr.，“Liu Hui and the First Golden Age of Chinese Mathematics，” *Mathematics Magazine*，vol. 71， no. 3 (1998)： 164. 经美国数学协会许可转载

在方块a中，第一行是空白，就像我们现在使用长除法时最上面一排会空白一样。第二行表示6538。第三行的9放在百位数列5的下方。注意方块b中，65除以9的计算结果是记入7。因此，在方块c中，取走6和5，并在百位数列代入余数2。整个计算过程，正如同用我们现在的印度-阿拉伯数字来做长除法一样。

利用这种算筹系统进行的算术基本运算与利用印度-阿拉伯系统进行的运算完全一样。公元前4世纪开始，中国的商人、科学家和旅行者就使用算筹，一直到16世纪，算盘取代了算筹系统。7世纪时，算袋是武官的标准配备。《孙子算经》明确详尽地说明了如何用算筹进行乘、除、开平方和开立方的计算。《孙子算经》中关于乘和除计算的说明，与阿尔-花拉子密的算术著作中那些关于印度-阿拉伯数字计算的说明相同。两种系统中对计算的描述几近一模一样，使得一些专家认为我们今日的印度-阿拉伯系统，可能是从中国传至印度的。《雪泥鸿爪溯数源》的两位作者如是说："这是唯一已知在概念上与印度-阿拉伯数字系统完全相同的数值系统。"

对多数古代文化来说，代表前三个数字的符号不是水平线就是垂直线，很可能是从手指或木棒的形状演化而来。到了代表四的符号时，通常不会看到四条垂直线或四条水平线，而是线条的组合，可能是四条线。对一些文化来说，在数字六之前，不会发生从并行线记号转换为其他组合的情况。中国的数字系统是最古老的系统之一，而我们可以在这个系统中发现一种手指计数或木棒计数的逻辑过程。代表六的符号不应该是六根垂直木棒，因为如果不数一数，很难分辨五根垂直木棒与六根垂直木棒——数字符号的要点是，不需要数。这与当代的计数符号非常类似，以四条垂直线和第五条水平穿过的线来作为5的记号。

小孩远在领悟颜色的意义之前，就知道了彩虹的颜色。数词（number words）与数概念（concept of number）的情况也一样。若被要求发明一个未曾见过的数系，你我可能想出一种希腊式或希伯来式系统。这样的发明很寻常（近乎自然），却像早期的台式计算机一样很难用。

远在任何人想到利用一种基底系统来书写数字之前，数字是写成记号，往往是五五成组。只要组数不是太大，其实不需要用个别的符号来代表数字。没有文字或符号来表示这样的"组数"的基数时，"不是太大"究竟是什么意思？这样的系统在如十个或二十个记号时没有问题，但个别数字没有名称或图画，这个系统在大计数时会失灵。

最近，我无意中听到两个孙女的一段对话。五岁的莱娜问十岁的堂姐苏菲说："为什么我的右手有五根手指？"苏菲的回答棒极了。

“这样我们才能正确数数。”除了小孩，谁能想出如此绝妙的本末倒置的答案？

在柏拉图的一篇短篇对话录中，一个雅典人极富智慧地问我们如何学会数数：

> 我们如何学会数数呢？我问你，我们如何发展出一与二的概念呢？宇宙的机制如何赋予我们人类了解这些概念的本能呢？还有许多其他的生物，它们天生的本能，不足以使它们发展出我们人类从天父那里学到的数数的能力。但是，以我们的情况来说，神在创造人类之初，便赋予我们理解世界面貌的能力，并且向我们展现他仍在不断展现的景致。

纯数学依赖“数”的意义。从几乎第一次接触数字开始，我们就能正确了解数字，而且在我们知道数字真正的意思之前就轻松自如地使用它们，这不是很了不起吗？那个雅典人论述道：

> 请回想我们十分公正的观察，若是数字从人类中禁绝，我们永远不可能有智慧。这是因为，如果一个生命没有理性的论述，他的灵魂就无法获得完整的德行。一个生命如果不能认得二与三，奇与偶，且对数字全然陌生的话，他就不能对事物做出理性的叙述，而只能对事物有感觉与回忆，虽然不会有任何事情可以阻止他拥有其余的德行，比如勇气与节制。

我们可以定义我们所谓的数是什么。但无论我们如何定义数，它的意义必须导向正常世界，那些与我们已经特意建立的原则一致的世界——和我孙女的看法一样，罗素的名言应该是正确的：“我们想拥有十根手指、两只眼睛和一个鼻子。”

---

[1] 这段引言出自《孙子算经》序，原文误认为出自《九章算术》。——译者注

[2] 一般认为，光绪二十五年（1899年），晚清官员、金石学家王懿荣发现甲骨上的甲骨文，这些甲骨文来自今日被称为“殷墟”的小屯村商代晚期遗址。——编者注

[3] 上一行为算筹纵式，下一行为算筹横式。用算筹表示数字时，从右边开始，先纵后横，纵横相间。此处作者将纵式、横式分别视为个位数与十位数，显然不对，因为纵式可能代表百位，而横式可能代表千位，等等。——译者注

[4] 陪葬的墓主年谱最后一年为西汉吕后二年（公元前186年），因此考古学家与历史学家断定这是《筭数书》抄写的年代下限。有关《筭数书》，可参考洪万生、林仓亿、苏惠

玉、苏俊鸿合著的《数之起源：中国数学史开章〈筭数书〉》，台湾商务印书馆，2006年。——译者注

[5] 此处作者所谓的矩阵系统（matrix system）意义未明。不过，如果它是指涉《九章算术》（以原注13内容为例），那么《九章算术》第八章的“方程术”，就是表现成矩阵形式的联立一次多元方程组之高斯消去解法。——译者注

第4章 印度送给世界的礼物

有些婆罗米（Brahmi）小数字（参见图4-1）在轮廓上与我们现代的小数字很相似。然而，婆罗米系统的概念非常不同。它不是一种基于十的乘幂的位值系统。相反，它比较像一种字母基底的数值系统，即使相当小的数字也需要用长串相连字符来表示。

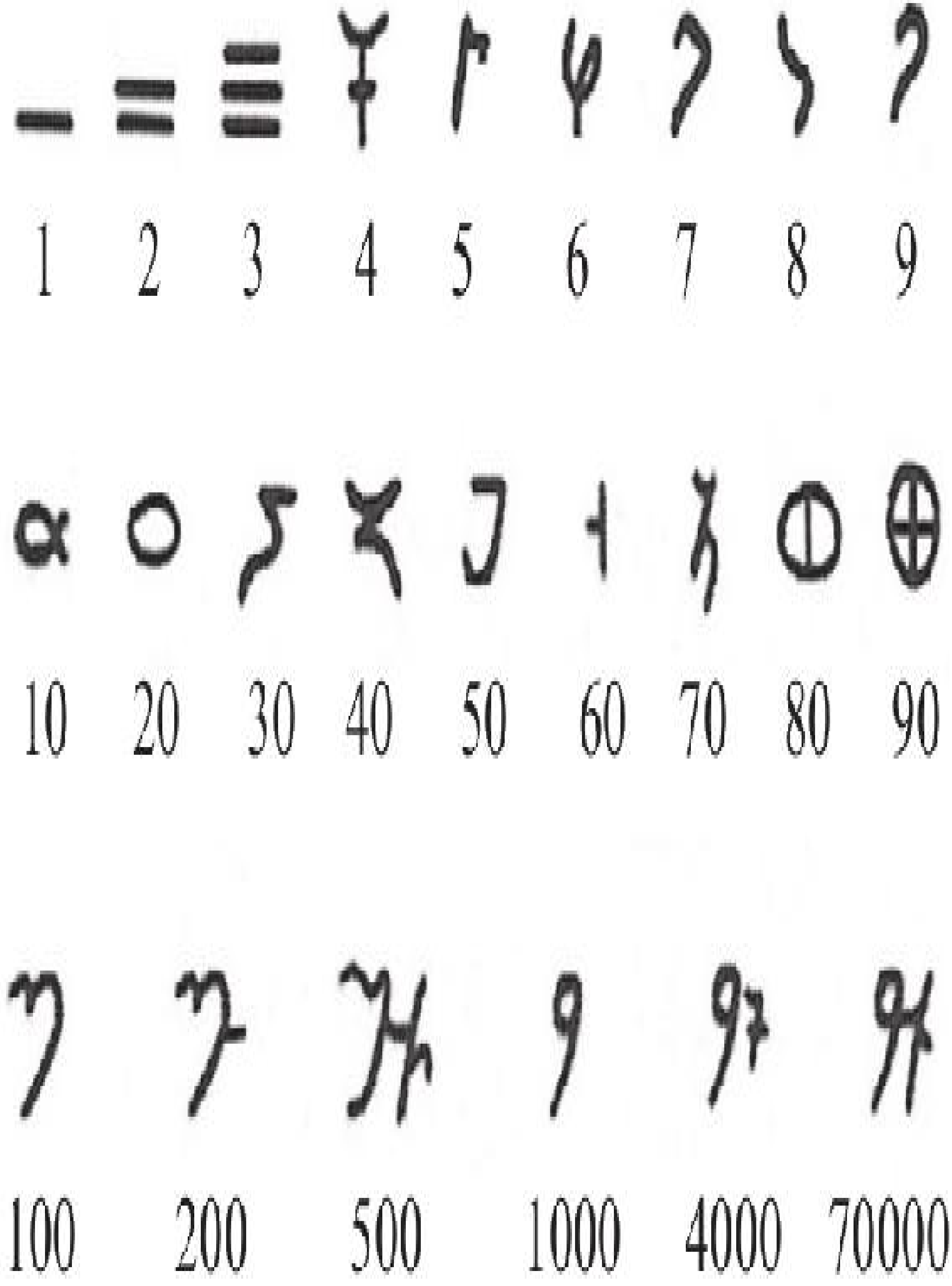

1 2 3 4 5 6 7 8 9
10 20 30 40 50 60 70 80 90
100 200 500 1000 4000 70000

图4-1 婆罗米数字

曾经有一段时间，人们推测该系统中用来表示4之后的数字的图案，来自首字母的形式，或是源于公元前3世纪婆罗米字母表的数词音节。但它们也可能来自更古老而无法追查的数值符号。一个较可能的来源是梵文的天城体书写体，这种文字刚开始是旁遮普的一种口语，后来扩展为“吠陀”（知识）的语言，这种书写媒介用来撰写通常为韵文形式或称为格言的短句形式的宗教赞美诗和祈祷文。数字对那些吠陀来说是不可或缺的，吠陀经常提到印度神祇的成就，他们摧毁了九十九座城市或赠送了六万匹马。有些吠陀文本记述了几组像是兆这种大数。后来吠陀被视为神圣的知识，包括对每日献祭的详尽天文时间测定的说明。而有些吠陀利用十的连续乘幂来描述大数字。

不幸的是，由于恶劣的亚热带气候，公元前第一个千禧年之前的印度数学遗产多半难以追溯。由于考古线索极少，印度数字的起源必须仰赖仅存的少数以石刻形式留存下来的记录。所以我们的数字如何演化发展的故事，仍然非常不确定。然而，那些使用了十进制位值数字的石刻碑文有一些提供了某种证据，说明古印度人很熟悉一种位值数字系统。

虽然这么说可能有点夸张，但检视梵文书写的数字之后，似乎可以猜想，那些数词的某些字母组合，很可能暗示了现代书写体在早期发展过程中的形状演化（参见图4-2）。

印度-阿拉伯数字	1	2	3	4	5	6	7	8	9	0
梵文	ekah	dvau	tryah	catvarah	pañca	sat	sapta	ashta	nava	suunyá
梵文符号	१	२	३	४	५	६	७	८	९	०

图4-2 印度-阿拉伯数字vs. 梵文

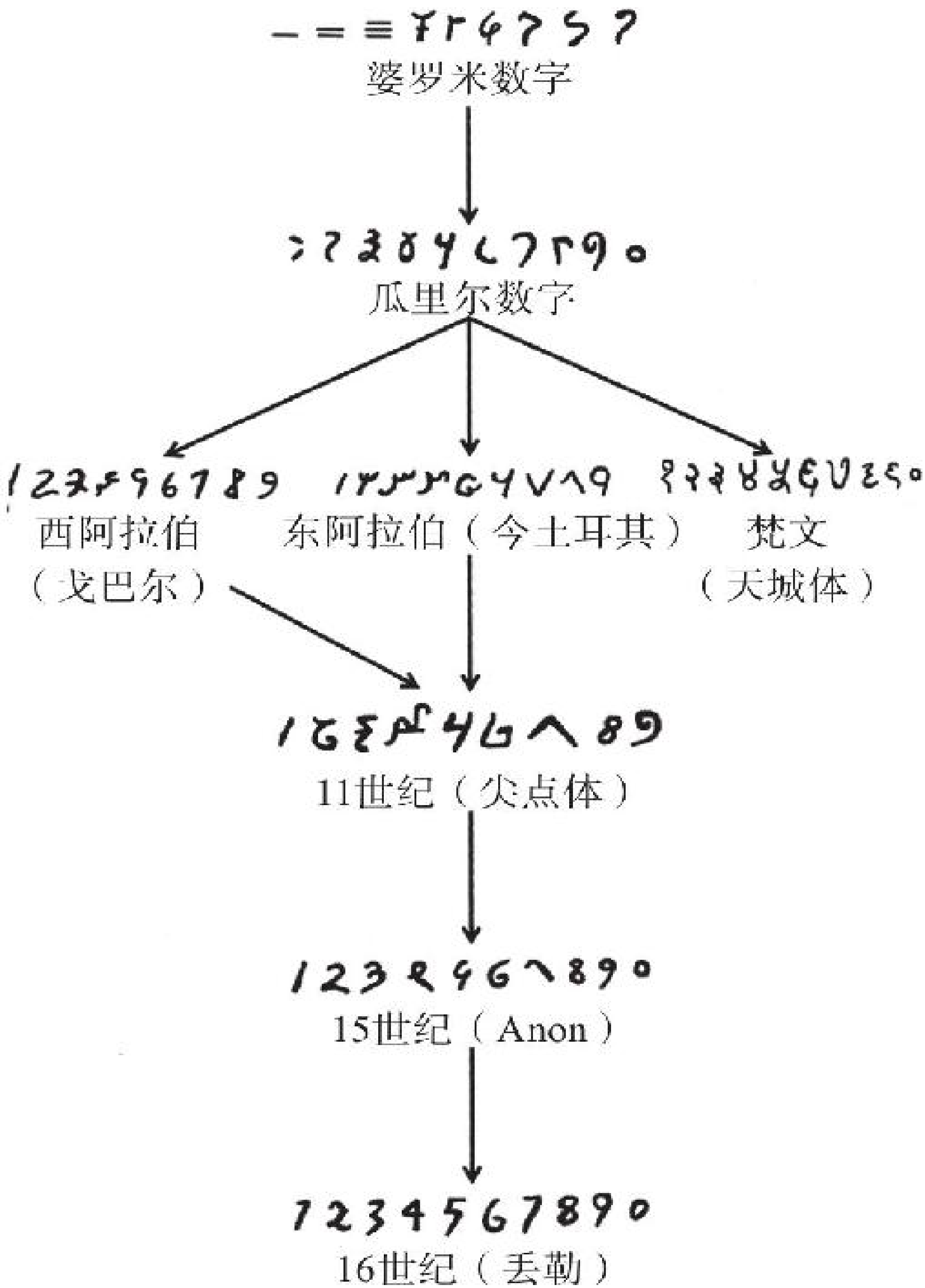

婆罗米数字
瓜里尔数字
西阿拉伯
（戈巴尔）
东阿拉伯（今土耳其）
梵文
（天城体）
11世纪（尖点体）
15世纪（Anon）
16世纪（丢勒）

图4-3 现代印度-阿拉伯数字演化表。重绘自Karl Menninger， *Number Words and Number Symbols：A Cultural History of Numbers*， trans. Paul Broneer (Cambridge， MA： MIT Press， 1969)，418.

这些数字让我们较全面地了解位值系统，以及一个将零视为数的系统。在位值系统中，数字依据它们彼此之间的相对位置而有不同的值。今日全球科学界已采用印度-阿拉伯系统。然而，在中东和远东使用的一些符号或多或少存在差异。东阿拉伯或印度语系的符号，在今日的巴基斯坦和伊朗使用。其他系统如日文，兼用印度-阿拉伯数字和汉字，阿拉伯数字横写，汉字直写。再者，汉字数字还有一种特别的书写形式，称为“大写”[1]，用于法律和财务文件，以防有心人添笔篡改，例如把二改成三。

图4-3表示我们现代的印度-阿拉伯数字的部分形态演变，这样的演化始于婆罗米数字。

历史往往难以明确界定，通常也不是线性发展。现代数字的形态演变，远比图4-3所示的过程更令人震惊。一个难解之谜是，从早期的书写体到今日的书写体，并不存在明确的血统关系。书写用的材料和工具，以及抄写过程中的错误，必定影响了这些数字的形状，使其与原始样貌大不相同。

我怀疑数字符号的设计和演化源于手指计数，不过没有任何确凿的证据。想象我们身处亚伯拉罕时代的乌尔城，要去市场买一条鱼。你可能会举起一根手指，表示你只想买一条鱼，或举起两根手指表示想买两条鱼。你的手的方向可能是垂直的或水平的。因此，代表“2”的符号可能以两根垂直的手指或两根水平的手指来表示，那些手指记号很快被描绘成水平的两杠，而随着时间流逝，为了能够更快速地书写，字体彻底变形，结果变得像现代使用的符号。

我们不知道婆罗米系统的真正起源，也不知道其他许多没有留下的历史记录的中间演化过程。公元前3世纪的婆罗米系统，是来自婆罗米字母、其他某种字母、古老的埃及数字、一种更早的印度河文化，或者来自更远古的数字吗？而瓜里尔数字这种位值系统，是否如同数学史家蓝丽蓉所宣称的来自中国？她认为到了1世纪，中国人已有了一种十进制位值系统，有九个记号和零的概念。

乔治·约瑟夫在其著作《孔雀之冠》（*The Crest of the Peacock*）中告诉我们，除了巴比伦人精巧的六十进制（以60为基底）位值系统之外，我们现代的位值系统完全来自印度。然而，罗伯特·卡普兰在他的著作《从零开始》（*The Nothing That Is: A Natural History of Zero*）中告诉我们，我们今日所用的系统是印度制，但它们起源于希腊。没有坚实的文献证据，无法填补这些历史空白。我们确实知道的是，不知何故，也不知何时何地，精巧的位值概念由印度人传给阿拉伯人，后来再传给欧洲人。

法国数学家拉普拉斯（Pierre Simon Laplace）曾经信心十足地说道：

> 印度给了我们能表示出所有数字的巧妙方法，利用十个符号，每个符号代表一个位值，以及一个绝对的值；一个深刻而重要的概念，以如此简单的形式出现在我们眼前，使得我们忽略了它的真正价值。但由于它的简易性和它适于所有计算的便利性，使得我们的算术成为最有用的发明；而当我们想起阿基米德和阿波罗尼斯（Apollonius）这两位古代最伟大人物的非凡才智都没能发现它时，我们必须更加赞赏这份显赫的成就。

东阿拉伯数字仍在埃及东边的阿拉伯国家使用，那里将这种数字称为印度数字。在摩洛哥，使用的是西阿拉伯数字（戈巴尔数字），这种数字称为阿拉伯数字。[2]尽管没有直接的证据，但尖点体数字与戈巴尔数字相似[3]，且似乎源于印度。戈巴尔数字来自印度，但尖点体仍是一团谜。20世纪早期，德国数学史家莫里兹·康托尔宣称，波埃修斯（Boethius）从戈巴尔数字创造了尖点体，并宣称早期印度数字传至亚历山大城的时间是在4世纪末两地的贸易终止前。[4]康托尔还声称，印度数字（没有零）在阿尔-花拉子密的《算术》（*Algorism*）被翻译成拉丁文之前，已传到欧洲超过一个世纪的时间，而它传入的时间是11世纪。

虽然从东方到西方有许多不同的形式，我们必定想知道，在它们从印度往四面八方传播的过程中，从一种文化到另一种文化，从一个国家到另一个国家，在超过一千五百年的时间里，它们的基本书写形式为什么几乎没有任何改变。关键是，虽然符号本身看起来或许不同，但婆罗米系统之外的每一个系统，都使用了位值来表示10的乘幂和零。婆罗米系统并非一种发展成熟的位值系统。它包含了代表10，20，30，40……90及100，200，300，400……1000的不同符号。婆罗

米人不会将两百二十二这个数字写成，它在位值系统中会表示成。这是因为是婆罗米系统中代表200的符号，则是代表20的符号。

问题仍未解决：西方世界这个带有零的数字系统是如何诞生的呢？为了解答这个问题，我们应该先回到手指计数、计算用沙板和算盘的时期。

在第二个千禧年前半叶，商人利用弯曲手指来计数和做简单的算术计算。商人举起他的手，手掌朝外，表示数字（参见图4-4）。

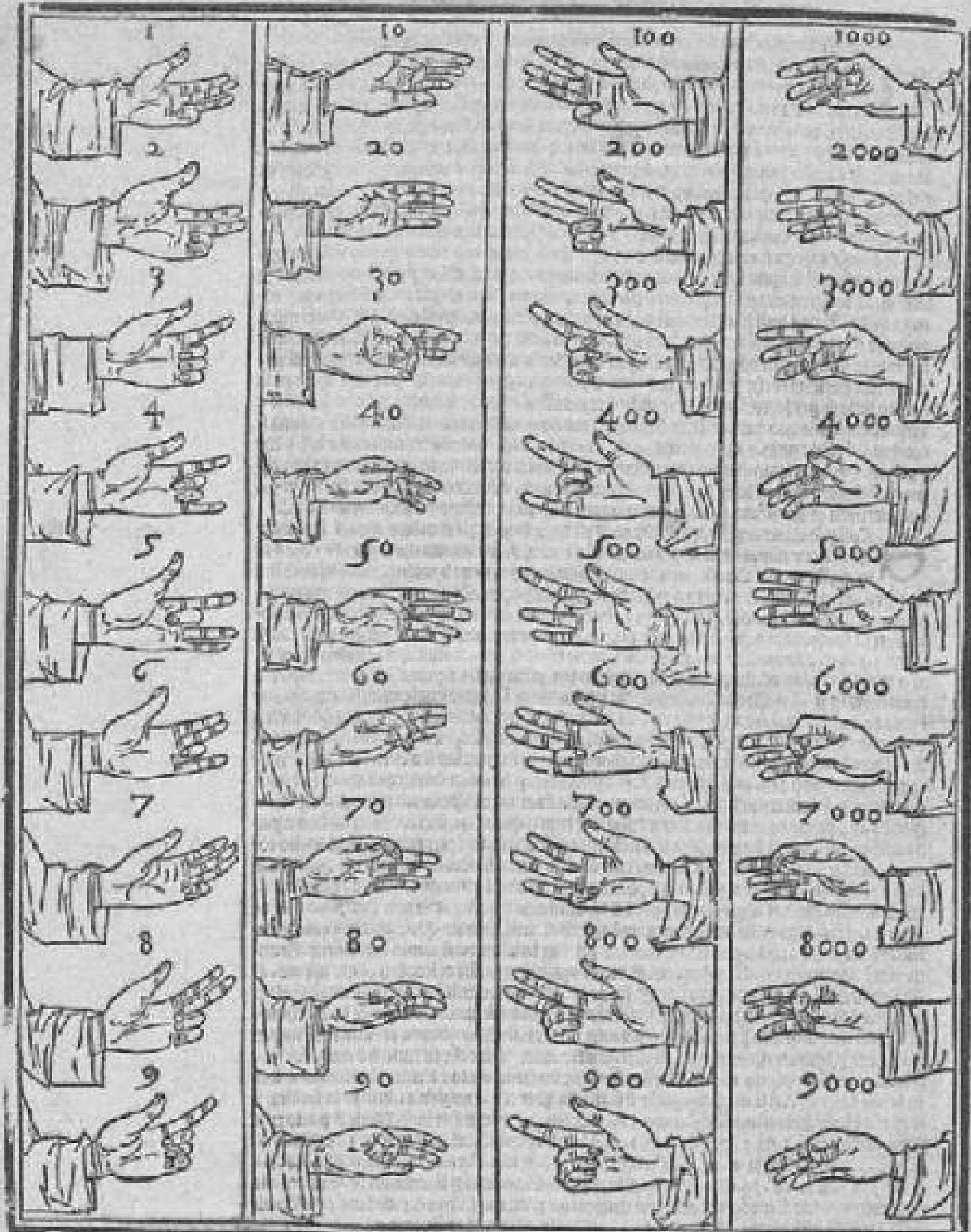

1
10
100
1000
2
20
200
2000
3
30
300
3000
4
40
400
4000
5
50
500
5000
6
60
600
6000
7
70
700
7000
8
80
800
8000
9
90
900
9000

图4-4 手指计数，引自帕乔利于1494年出版的《大全》（*Summa de Arithmetica*）一书

如图4-4左边第一列所示：

要表示数字1，只把第五根手指半弯；

要表示数字2，只把第四根和第五根手指半弯；

要表示数字3，只把第三根、第四根和第五根手指半弯；

要表示数字4，只把第三根和第四根手指半弯；

要表示数字5，只把第三根手指半弯；

要表示数字6，只把第四根手指半弯；

要表示数字7，只把第五根手指全弯；

要表示数字8，只把第四根和第五根手指全弯；

要表示数字9，只把第三根、第四根和第五根手指全弯。

仍然是使用左手，以一组不同的符号来表示从10到90的数字。举例来说，要表示10，将食指指尖置于拇指的底部，让手指的形状看起来像希腊字母δ，它代表的是10这个数字。

这种手势符号不过是语言不通的商人之间的一种表示数字的肢体语言，因为它并没有发展出算术计算方法。纽约商品交易所、美国证券交易所和其他证交所仍使用这种手势信号，“公开喊价”时手势信号代表买卖的指令：交易者举起手指，手掌向内表示要买，手掌向外表示要卖。利用这种复杂的手势，可以表示各种各样的交易。

这让我想起大概五十年前在委内瑞拉奥里诺科河的卡布鲁塔的一次冒险之旅。一个市集日早上，我早早醒来，发现大家都到村子广场喝咖啡。帕内里人，那个地区的原住民，贩卖着鹦鹉、猴子、虎猫和淡水豚。我学到帕内里语中用来表示身体部位的词可用以代表数字。表示手的词代表五；表示“另一只手”的词代表六；表示“两只手”

的词代表十。而其他的身体部位，诸如“脚”“另一只脚”和“两只脚”则代表更大的数字——我记不清了，但我想是代表11，16和20。

人类发展出加和乘的能力，必定始于某种标记体系，无论是来自手指计数、小石子计数还是某种更富想象力的东西。计数发展的早期阶段应该是通过具体的方式，逐一地指向对象。在阿兹特克人遗留的语言里，他们使用了一块石头、两块石头、三块石头等的数字。南太平洋语言则是数一个水果、两个水果和三个水果。然而，随着时间推移，计数——利用诸如手指、石头、水果和谷物这类具体对象——发展进入抽象的阶段，此时计数的对象的特性已不再重要。这就是数学。借由重复数手指或某种其他标记体系，形成数的抽象概念。

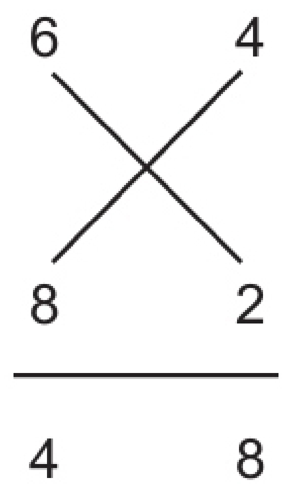
我们有强有力的证据证明，所有的数系都是从数手指、脚趾和身体其他部位演化而来。儿童数数时自然地使用手指，与数字一对一对应。对算术的发展来说，或许这是至关重要的过程。

生活在新几内亚偏远高地的原住民部落亚浦诺，使用一种可以计数到三十三的精巧系统。这种系统依某种特定的次序来数每根手指，接着是以身体部位来计数，交替地从一侧到另一侧。亚浦诺计数系统有一项明确的优点。美国小孩数手指时，是从拳头开始一根一根伸出手指，直到数完为止。最后，他们所伸出的手指头数量便是答案。它没有预先确定次序，小孩在计数时可以从任何一根手指开始，随意伸出其余任何一根未伸出的手指，即便其中存在某些文化标准。由于亚浦诺系统需要以明确的次序来计数，它的优点是最终的答案仅与最后数到的那个身体部位有关。

回到过去那个在市集中书写仍不太方便、普遍采用手指来计算的时代。美国数学家大卫·尤金·史密斯（David Eugene Smith）根据20世纪20年代发现的历史证据写道：“数字记法的一般目的，是当彼此不了解对方的语言时，有助于大笔国际交易的谈判过程，记住算盘计算的数字，并执行简易计算。”现存关于古代手指计数的唯一完整记录是《利用手指计算与谈话》（*De computo vel loquela digitorum*）抄本。8世纪本笃会僧侣圣比德（Venerable Bede）撰写了这部著作，他在中世纪学者中以计算每年变化的复活节日期而闻名，这个日期被设计为绝对不与犹太人的逾越节同一天。由于所有其他的教会节日都依复活节而定，圣比德的计算更显重要。圣比德说明了如何只利用屈伸手指，来表示1到100万之间的所有数字。从手指记法发展到手指计数，并进一步发展到手指计算。的确，将两个数相乘

时，我们并不需要知道5×10以上的乘法表。小数字相乘可以简化成数手指，乘上10并加上100。举例来说，将6乘以8时，两个数各减去5，得到1和3。左手伸出一根手指，右手伸出三根。数伸出的手指（1＋3＝4）并乘上10，得到40。现在，将两手弯曲的手指相乘（4×2＝8）并加上前述的40，得到48。[5]要将11至15之间的两个数相乘，两个数各减去10；以伸出的手指表示减完后的两个数。数双手伸出的手指并乘上10。将此结果加上两手伸出手指数的乘积，然后加上100。举例来说，要将12乘以14，两个数各减去10，得到2和4。左手伸出两根手指，右手伸出四根。数伸出的手指（2＋4＝6）并乘上10，得到60。将两手伸出的手指数相乘（2×4＝8）。将100，60和8相加，得到168。[6]

16世纪的文本显示，当笔算行得通时，这种简单的乘法如何进行。它们甚至指出乘法符号的起源。如果想将6和8相乘，从10各减去6和8，得到4和2。此时，如下图所示，将这四个数写在正方形格子上。为了得到答案48，从6减去2得到4，放在十位数那列。接着，将右列的两个数相乘，得到8。



英国神经心理学家布莱恩·巴特沃斯在《大脑如何与生俱来就能做数学？》（*What Counts: How Every Brain Is Hardwired for Math*）一书中提出一个问题：为什么左顶叶（大脑中掌管手指活动的区域）也是大脑中掌管计算的区域？移动手指之于计数，是否如同眼睛之于看的必要性？若是如此，巴特沃斯提出的问题“计算能力是否源于我们用手指来计算”便有了答案。他推测这是对的。为了得出答案，我们需要将更多手指谜题放在一起思考。加拿大神经外科医师韦

尔德·潘菲尔德著名的运动皮质（大脑控制身体运动功能的部分）图显示，控制身体相邻部位的细胞，在运动皮质中也是相邻的。

但不仅如此。需要做更复杂移动的身体部位，占据了大脑大范围区域。身体中较小的部位需要更错综复杂的移动，例如手指。相较于如手臂等较少错综复杂移动的身体部位，手指在运动皮质中占据更大的区域。另一项重要的考虑，来自关于阅读盲文（即用手指阅读）的人的极其惊人的研究结果。他们掌管手指运动的运动皮质，占据了大脑中更大的区域。弹钢琴的人的运动皮质，是否出现同样的现象？那么法庭速记员呢？

所以，我孙女或许终究是对的：我们每只手有五根手指，因为“这样我们才能正确数数”。

手指计数的原理延续到卵石标记的方式，继而很可能带来沙板计算和算盘的发明。我说“很可能”，是因为除了一些晚期的历史传说之外，没有可信的证据支持这样的说法。但这是值得深思的想法。数一百颗卵石，比数一百头四散放牧的绵羊容易得多。而数十堆卵石，每一堆有十颗，也比数一百颗卵石容易得多。埃及人、希腊人和中国人利用不同大小的卵石，运用这类卵石计数的技巧进行日常计算。同样大小的每一颗卵石，代表一堆较小尺寸的卵石，例如十颗卵石表示的是一百颗更小的卵石。后来这个系统演化为一种不需区分大小的形态，因为卵石计数者学会利用不同的位置，来放置代表十的卵石和代表一的卵石。

这看起来或许并非极为先进的思想，但它促进了算盘概念的发展。早期的算盘仅利用了卵石计数体系，当中的卵石是直线排列的：一条线代表一，一条线代表十，一条线代表百……一堆有四百二十三颗的卵石可能不容易数，但如果每一颗大卵石代表一百，每一颗中卵石代表十，每一颗小卵石代表一，那么四颗大卵石、两颗中卵石和三颗小卵石，就代表423（参见图4-5）。

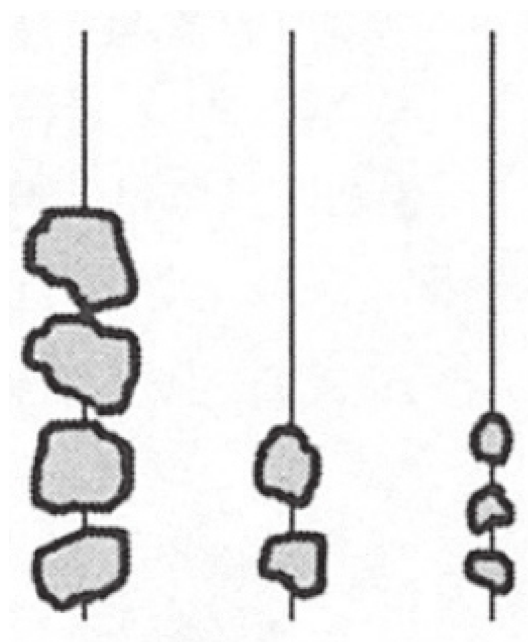


图4-5 画在沙上用卵石计数的线。资料来源：Georges Ifrah， *The Universal History ofComputing： From the Abacus to the Quantum Computer* (New York： John Wiley & Sons，2001)， 11.

沙板看起来更像是小型的沙盒，之所以使用沙板，是因为这种方法在计算时需要移动数字，并且擦掉计算过程。这种沙板就像百年前语法学生所用的石板和板擦，又或者现在所用的白板，计算过程中会在板子上写下、移动或擦掉数字。

计算板可远溯至巴比伦时代。然而，除了几件来自希腊的样本，我们没有实际的样品。1846年，希腊萨拉米斯岛上发现了一块白色大理石平板（现藏于雅典国家考古博物馆），板上刻有平行列。它为我们提供了直接的证据，证明使用卵石的计算板，至少可回溯至公元前300年。借由大流士花瓶这项间接证据，我们也证明了计算板至少可远溯至公元前4世纪（参见图4-6）。

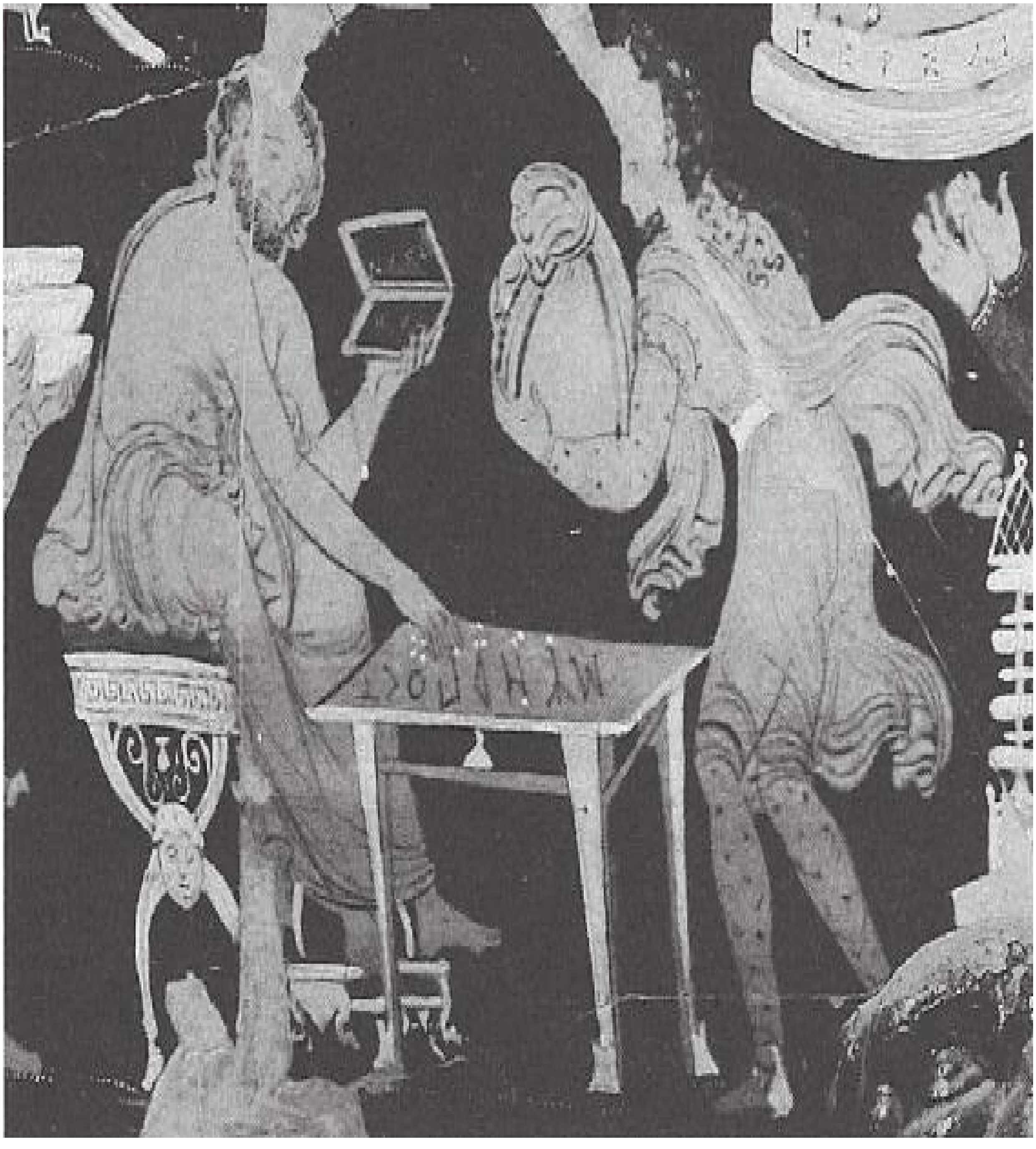


图4-6 大流士宫廷的王室司库和他的计算板，公元前4世纪。图中站着的男人带来了战利品，这里的算子未成列放置，而是直接放在代表数值排的符号下方。司库的左手拿着某个看起来像iPad的东西，那其实是一个记录算盘上依序求得的数量的垫子。资料来源：大流士花瓶局部（国家博物馆，那不勒斯）

从巴比伦人、希腊人和罗马人的计算板来看，数字表示法遵循了位值法则。他们没有代表零的符号；没有任何一个计算板需要，因为

空的一列可以表示有一个位阶没有数值。这个概念带来接下来的发展：将珠子穿在直线上，亦即更现代化的算盘。

罗马算盘上有金属球滑动槽。[7]图4-7的算盘上有八个十进制值，（从右到左）标记着I，X，C，∞……分别对应个、十、百、千……一直到百万。暂时略过图4-7右侧的前两个滑动槽。[8]标记上面是单个金属球，或在各个槽上滑动的算子。如果要表示小于5的数字，算盘师傅就将相应数量的算子，往字母标记移动。如果要表示5至9之间的数，算盘师傅会先将对应那个数的单个算子，往上移动至表示5之处，接着往上移动那一列的其他算子，直到表示出所要的数。所有小于一千万的数字，都可以用这种方式表示。举例来说，图4-8说明了5372这个数字的表示方法。

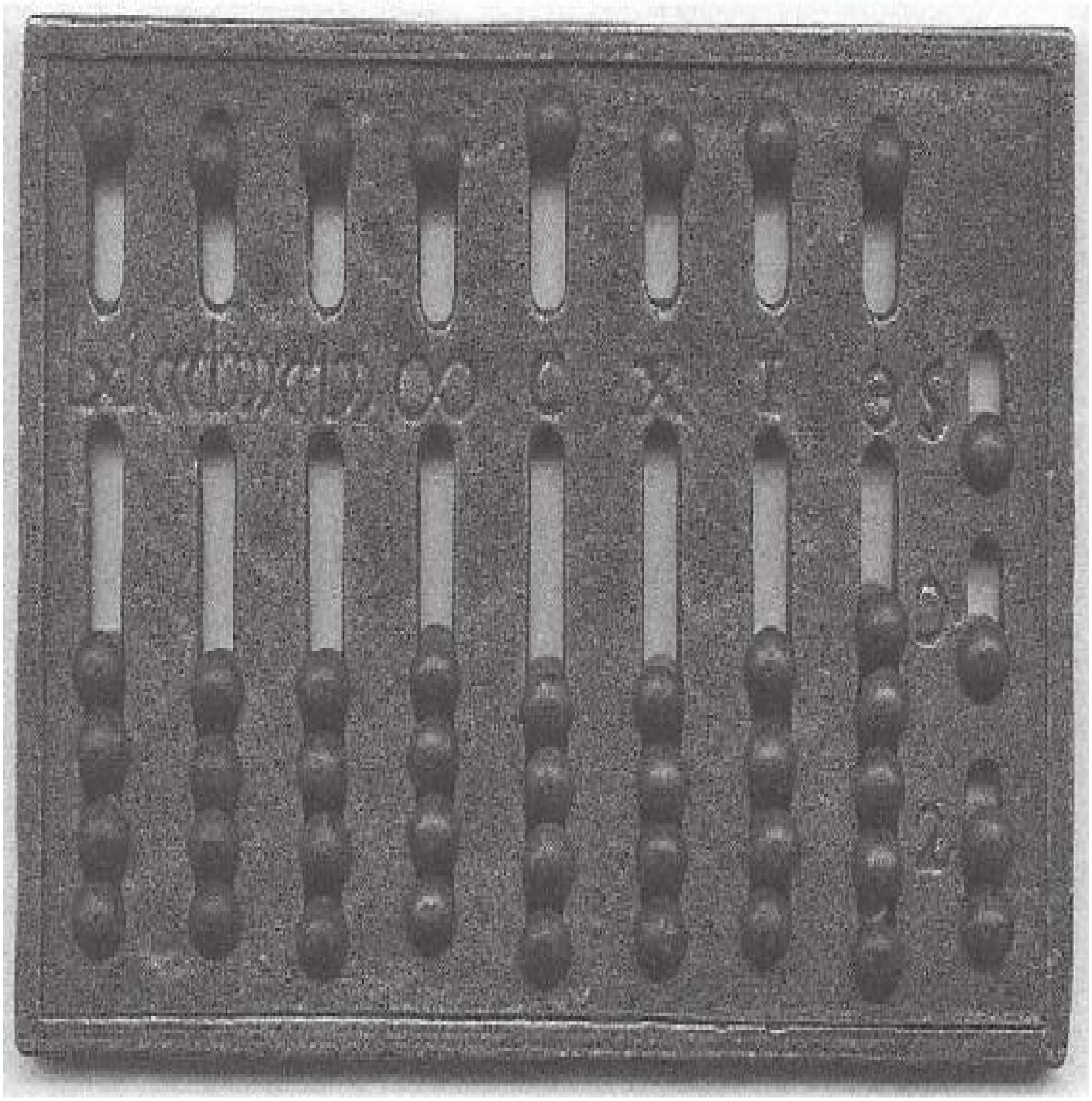


图4-7 罗马算盘的现代复制品

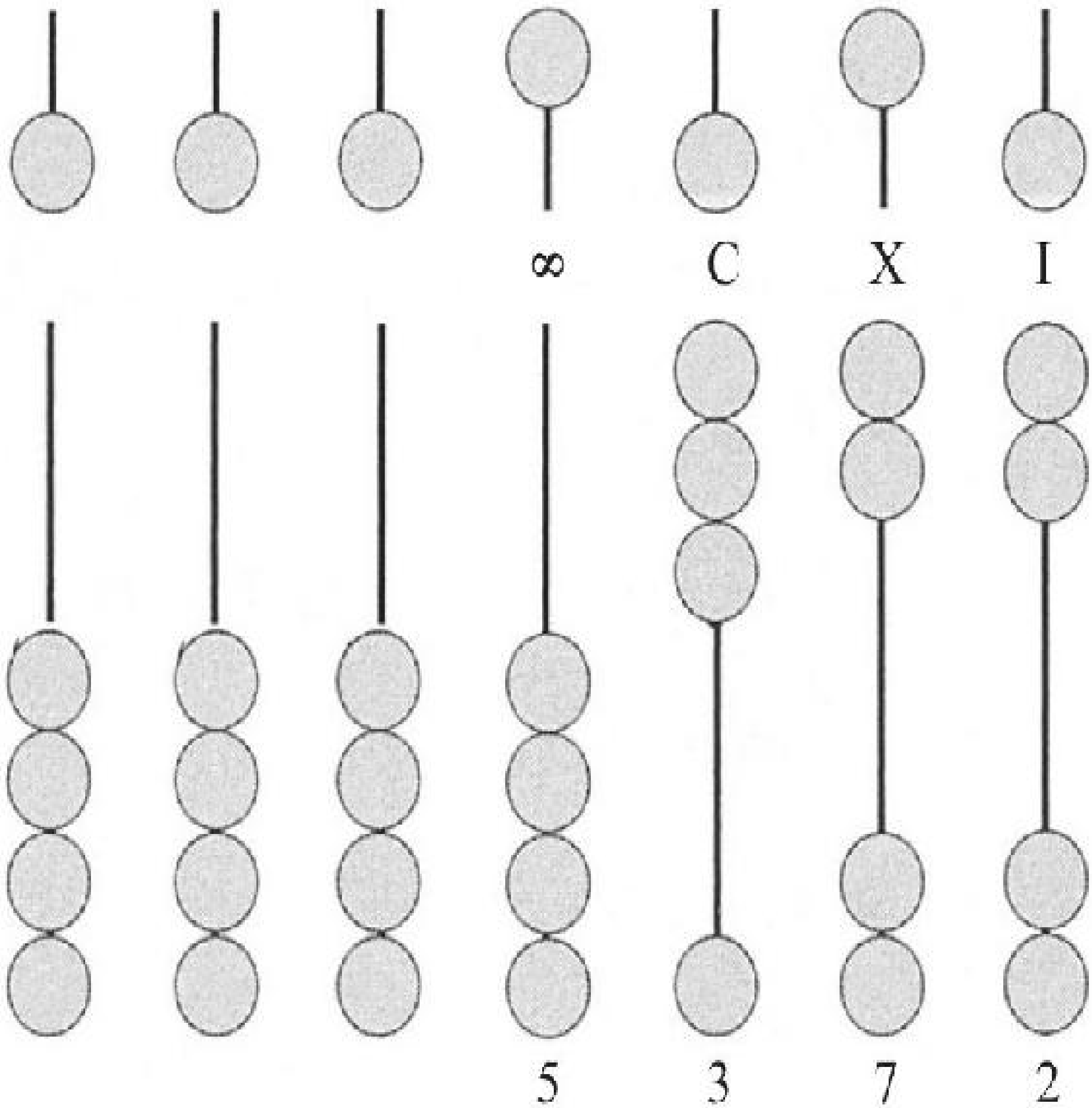


图4-8 数字5372在算盘上的表示方法

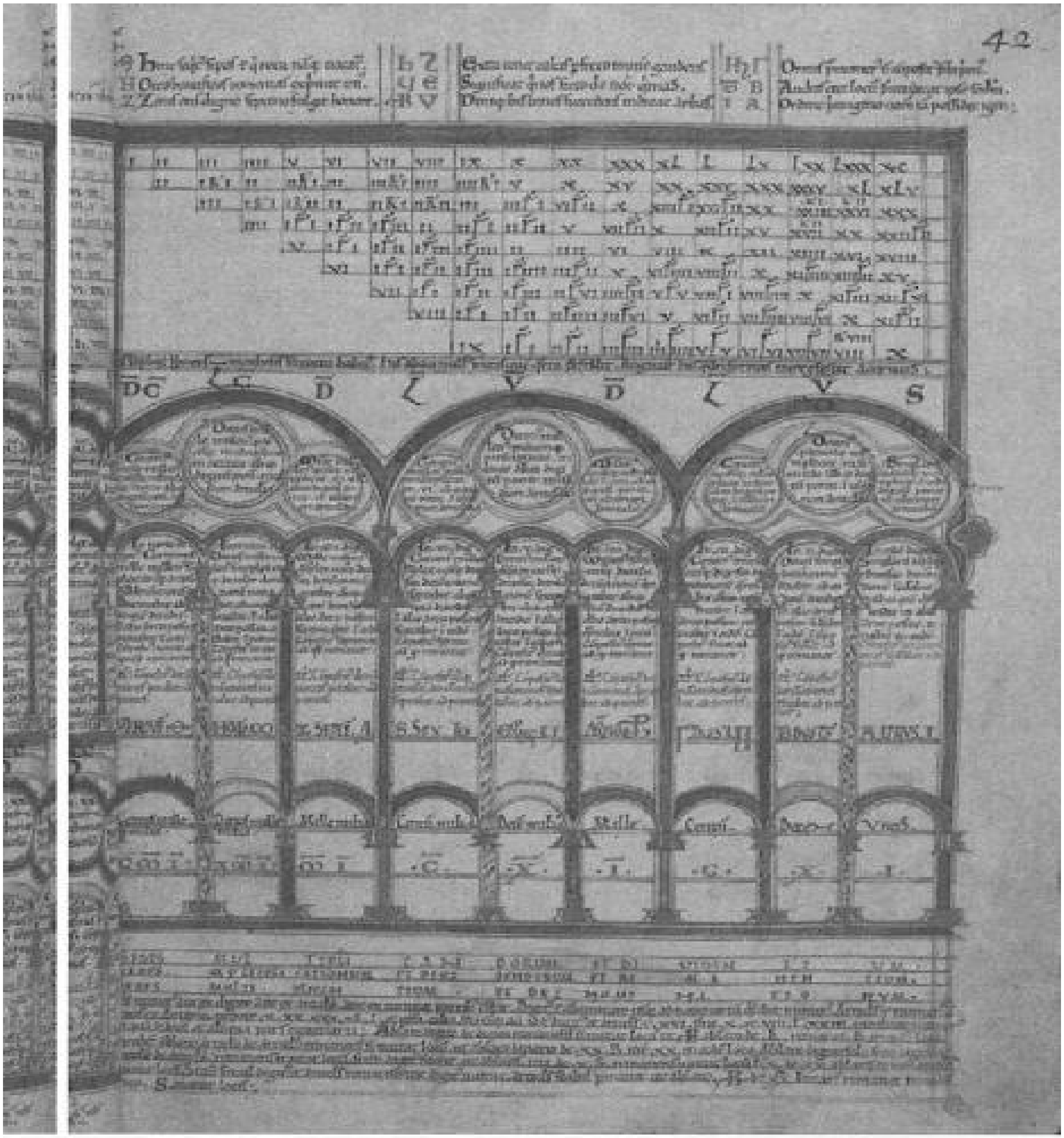
显然受到这种算盘的启发，10世纪时西方的数字会写成依数字次序排列的罗马字母。例如，5372会写成V.III.VII.II，模仿图4-7的算盘中标记∞，C，X和I的四个槽里的算子。

接着是吉尔伯特的计算板。吉尔伯特于950年出生在法国中南部的奥弗涅，在欧里亚克圣哲拉的修道院接受教育。967年，他离开修道院，前往阿拉伯统治下的西班牙旅行，在那里的三年间，他研读数

学、学习阿拉伯语，并熟悉了印度数字。他回到兰斯后，获得教堂学校的职位，教授数学和算盘计算。他的职业历程相当有趣——教学，接下来的职位是：先担任大修道院院长，接着是大主教、神圣罗马帝国皇帝奥托三世之子的家庭教师、教宗顾问，最后以非常年轻的四十九岁之龄成为教宗。他的原名是Gerbert d’Aurillac，在迈入新千禧年之前的厄兆年，他成为教宗西尔维斯特二世（Pope Sylvester II）。

10世纪晚期至12世纪中叶，吉尔伯特算盘一度广受欢迎。我们对原始的吉尔伯特算盘所知甚少（没有任何吉尔伯特算盘留存至今），但一些新近发现的手稿被认为是吉尔伯特算盘的具体范例。最近在东卢森堡的埃希特纳赫的本笃会修道院发现的埃希特纳赫手稿（*Echternach* manuscript， 约1000年）便是一例，而在英格兰剑桥的托尼修道院发现的“复活节计算表”手稿（*Computus* manuscript，约1110年）则是另一个例子。

检视这份“复活节计算表”的一页，我们发现一块计算板的各列是依照10的乘幂的等级来排列的。在图4-9中，右数第六列顶端，在$\overline{\mathrm{C}}\ (=100000)$下方，我们发现一个单一的“算子”表示5（在真实的中世纪计算板上，那是一个看起来很像y的符号）。吉尔伯特计算板上的原始算子是以兽角尖刻成，可能因为每个算子类似有顶点（apex）的圆锥形状，所以称为“尖点体”（apices）。[9]那些尖点（体）标记着奇怪的符号，看起来很像我们今日的印度数字（如图4-3印度-阿拉伯数字演化表从上往下的第四层）。虽然雕刻的兽角只是出于美学设计，在算术上没有特别的意义，吉尔伯特仍热切地在兽角上打造出无数尖点，而在他之后的其他算盘工匠制作出他们自己的算盘，他们所用的罗马算盘算子材料包括象牙、金属或玻璃。


42

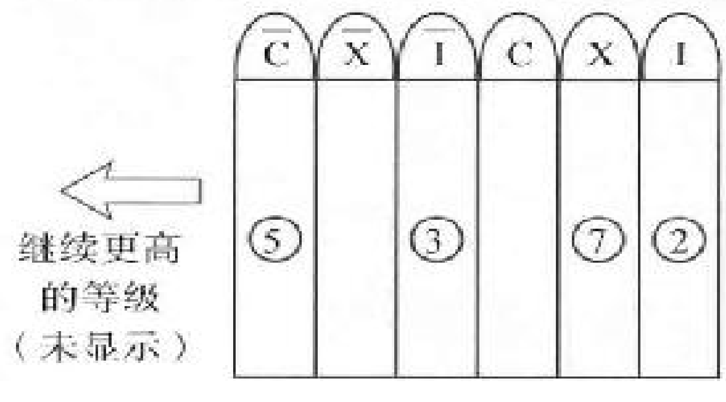

C̄
X̄
Ī
C
X
I
5
3
7
2
继续更高的等级（未显示）

图4-9 吉尔伯特计算板上显示的尖点5，3，7和2，代表的乘幂依序是$\overline{\mathrm{C}}$ $(=10^6)$、$\overline{\mathrm{I}}$ $(=10^4)$、X$(=10^2)$、I$(=10^1)$。Computus manuscript. Written at Thorney Abbey, Cambridgeshire, ca. AD 1110.Oxford, St. John’s College MS 17 Fols. 42r. 资料来源：http：//digital.library.mcgill.ca/ms-17/folio.php?p=42r&showitem=42r_8Math_1cHinduNumerals#. 经牛津大学圣约翰学院院长及其他成员许可复制

用一个带有符号标记的单一对象取代一组特定的卵石，这个想法很新颖；那些尖点体在西方首次出现。对于利用印度-阿拉伯数字来思考的我们而言，吉尔伯特计算板或许看来像是从罗马计算板自然发展而成的产物，但在当时，这是符号发展史中罕见的一次大飞跃。吉尔伯特想必听闻过阿拉伯构想那种令人惊叹的可能性。他应该已经意识到，利用新的数系进行笔算所带来的优势，但他的计算板仅作为计算工具，没有笔算的需求。因为西方未确切理解新数字的真正意义，吉尔伯特的门徒持续使用那些神秘的符号，并未注意到它们真正的潜在力量。

那些相同的符号展现了多样化的形式，有时旋转不同的角度，就好像毫不在意它是否正确地朝上或朝下。这可能是这种计算板的本质所造成的结果，计算板没有严格的方向规定，也没有任何既定的正确用法——一位算盘师傅与一个从计算板另一边看的人，所看到的可能不尽相同。在图4-9（“复活节计算表”手稿）中，数字3和8在同一份手稿里可以看出有两种不同的书写方式。但无论写成哪一种形式，整个系统是相同的，都只有九个符号，每个数都可以表示出来，而且每个数都可以描绘出来，不需要将画笔从计算板上移开。

10世纪至12世纪间，计算板是一张羊皮纸或一块开槽的板，其中垂直的列依10的乘幂排序，这种计算板是西欧学习实用算术的主要方法。吉尔伯特算盘的程序称为“算法”，用鹅毛笔和羊皮纸来进行模拟。

---

[1] 即壹、贰、叁、拾等。——编者注

[2] “戈巴尔”是对西阿拉伯数字的称呼。Gobar这个词源自阿拉伯文ghubar，意为“尘土”。它指古代在泥土或沙土上做计算的形式。

[3] “尖点体”的原文是apices（apex的复数形式），这是个极含混的词。——译者注

[4] 值得注意的是，在这几段中引述的权威学者的看法，多是从大量不同的来源汲取信息，其中很少是今日史学家认为可靠的资料。

[5] 这行得通，因为$ab=[(a-5)+(b-5)]\times10+(10-a)\times(10-b)$。

[6] 此式成立的依据，来自代数恒等式$ab=[(a-10)+(b-10)]\times10+(a-10)\times(b-10)+100$。一个类似的方法对任意两数也行得通，但会变得更复杂，因为两只手都可能必须伸出超过五根手指。这个想法是利用公式$ab=[(a-c)+(b-c)]\times c+(a-c)\times(b-c)+c^2$，其中$c$是数字应该化简的量，以便将乘法运算降到一个可处理的数字上。不幸的是，较大的数字需要知道如何求c的平方。

[7] 已知的罗马算盘只有两个，一个在巴黎的纪念章陈列馆（Cabinet des Médailles，法国国家图书馆），另一个在罗马的戴克里先浴场博物馆（Museo delle Terme Diocleziano）。

[8] 右数第二个槽里有五个算子，表示一盎司的倍数，每一个算子代表一盎司。在较短的槽里的算子代表六盎司。右边第一个槽分成三部分，表示半盎司、四分之一盎司和三分之一盎司。

[9] apices这个词在中世纪盛期的计算文本中出现时，意义含混不清。它有时指吉尔伯特算盘本身，也被称为毕达哥拉斯弧（arcs of Pythagoras，尽管毕达哥拉斯当时没提过这类想法）。有时它更恰当地意指算子本身，算子是圆锥体，所以有尖点。还有时它指的是出现在算子上的符号。在本书接下来的篇章中，它指的是实体算子本身。

第5章 符号在欧洲的启蒙趣事

将1＋1简记为2，这使2成为一个新颖且“随意决定”的符号。“随意决定”指的是，使用其他任何符号同样可运作良好。用2来代表1＋1，而不是用＜，3，∞或其他任何符号，是因为某些印度人选择使用它。

——德·摩根（Augustus De Morgan）

说来奇怪，我们奇妙的现行数系传入欧洲至今，仅仅用了几个世纪的时间。这一切是否主要该归功于比萨的李奥纳多，仍存在争议。他是那个时代最伟大的数学家之一，他的名气主要来自著名的兔子繁衍问题，而我们更熟悉他的另一个名字“斐波那契”（Fibonacci）。可以确定的是，他并不是发现如何解答兔子问题的人，大约在迈向第一个千禧年之际，古印度提出这个问题，用以描述梵文诗歌的诗体结构和蕴藏的韵律。

斐波那契留下了一幅画像（参见图5-1）。这幅画绘于13世纪中叶，他看起来像个很好相处的男孩，大眼、小嘴，仿佛在探究引人好奇的事物。斐波那契年轻时和父亲周游地中海，在埃及、叙利亚、希腊和普罗旺斯遇见了神职人员、学者和商人。他学习到在交易中使用的数系。回到比萨后，他于1202年撰写了《计算书》，1228年重新修订。《计算书》是一本讨论如何不用算盘来计算的著作，用以说服西方商人，阿拉伯数系的计算方式优于当时惯用的罗马系统。这不是西方第一本叙述阿拉伯数字的书。“维希拉努斯抄本”是西方最早记有阿拉伯数字的手稿，当时已有两百五十年历史（参见图5-2）。而阿尔-花拉子密《算术》的拉丁译本在12世纪时已出现。

图5-1 斐波那契

然而，《计算书》在印刷术发明之前约两百五十年便已出现，那个时代没有地方性公共图书馆，知识是口头传播的。在意大利，印度-阿拉伯数字很早就传入，商业算术的实践者，也就是计算师傅，在担任算术、几何和代数的私人及公共教师时传播这项技术，并撰写相关的二流著作，展示重要的知识。他们抄袭其他手稿，经常改变问题的数值或调整问题来掩饰来源。在《计算书》首次出现约七十五年后，来自全欧洲的学生（波希米亚人、波兰人、法兰西人、日耳曼人、西班牙人）造访威尼斯和比萨等学术中心，传播了关于阿拉伯数字的信息。

Scire debemus in Indos subtilissimum ingenium habere et ceteras

gentes eis in arithmetica et geometria et ceteris liberalibus

disciplinis concedere. Et hoc manifestum est in nobem

figuris quibus designant unumquemque gradum

cuiuslibet gradus quarum hec sunt forma

9 8 7 6 5 4 3 2 1

图5-2 维希拉努斯抄本

17世纪晚期之前，大学里通常不教授代数。部分原因是，当时的大学是训练神职人员的主要场所，或培养医生和律师。反之，数学在*bottegas*（计算工作坊）蓬勃发展，这是14世纪和15世纪北意大利的一些计算学校，计算师傅在这里以方言教授商人和艺术家商业算术。“计算学校”（abacus school）或“计算传统”（abacus tradition）这类名称，源于这些学校的学生是以斐波那契《计算书》的方式学习数学，而非以算盘作为计算工具。abbacist（计算师）一词，用以称呼善于以印度-阿拉伯数字进行计算的人，区别于用算盘来计算的人。14世纪中叶，仅佛罗伦萨一地，就有至少一千两百名学生在该城约二十所计算工作坊就学。这些计算师傅以相当吸引人的方式，编写附有图示的专著，收录数百则算术和代数问题的解答，那些问题的难度往往超出学校所教授的课程。

查理曼大帝几乎征服了整个西欧，在公元800年成为神圣罗马帝国皇帝，他了解到欧洲在科学和医学方面远远落后于阿拉伯国家。他下令他的王国中每一所大教堂和修道院都必须开设学校，进行大众教育。那些学校除了教授几何和算术之外，几乎不教其他任何数学或科学。查理曼大帝逝世后，整个课程着重于拉丁文、音乐和神学。然而，多得出奇的富有天赋的教师和好奇心旺盛的学生将中世纪的课程，提升至博雅教育的层次：先是三艺（*trivium*），包括语法、逻辑和修辞；接着是四艺（*quadrivium*），包括算术、几何、音乐和天文学。任何顺利通过三艺的人，就会被视为博学。

到了12世纪，正当行会（guild）开始形成之际，最早的大学也设立了。这些大学通常是由师生自发建立并组织，自然独立于大教堂和修道院的学校之外。当然，因为当时担任圣职的人员是唯一受过教育者，大学教师都与教会有关联。

大学里的学生都只是儿童，通常不到十二岁。他们会花四年时间学习拉丁语法，若顺利通过，会被授予语法硕士学位。文学学士学位需要顺利完成三艺，这意味着更高阶、更受尊敬。文学硕士学位需要再花三年顺利完成四艺，这是当时可以取得的最高学位，非常难达到。要担任教职必须有文学硕士学位，但报酬微薄。

有很长一段时间，斐波那契的《计算书》是关于计算方法（abacus methods）的唯一已知包罗广泛的来源，或许因此被视为将

印度-阿拉伯数字传入西方的媒介。[1]现在有一些普及著作宣称是斐波那契将印度-阿拉伯数字引入欧洲。然而，20世纪60年代以来，几位史学家认为，在斐波那契时代前后一些涉及印度-阿拉伯数字计算的书籍并未提到《计算书》。而更近期，2002年，丹麦数学史家延斯·贺鲁伯认为，关于计算的书籍是从伊比利亚和普罗旺斯传到北意大利的，指出将印度-阿拉伯数字引入欧洲的人并不是斐波那契。

斐波那契在《计算书》的序言中写道：

> 当我的父亲在远离我们家乡的贝贾亚（Béjaïa）海关——为经常聚集在那里的比萨商人设立的——担任行政官员时，他将年幼的我带到他身边，希望为我寻求一个有益而安逸的未来；他希望我在那里学习数学并且受教一段时间。在那里学到了令人惊奇的九个印度数字的技术……

这里只提到九个印度数字，所以不包括0，至少没有把它当作一个数字。

贺鲁伯宣称，到了12世纪早期，印度-阿拉伯数字已经由伊比利亚和普罗旺斯传入拉丁文化。但20世纪的史学家将引入印度-阿拉伯数字归功于斐波那契，是因为在他那个时代，意大利的商业教学应用的仍是罗马数字。之所以造成混淆，似乎是因为斐波那契的著作一直使用阿拉伯数字，来讨论大家熟悉的数学。

贺鲁伯进一步告诉我们，意大利关于计算的代数并非由斐波那契所启发，而是来自意大利以外，而且意大利商人对计算传统中所教授的这类内容早已有迫切需求。他又提到：“我们可以从分析中得知，13世纪开始的计算传统是**非斐波那契式传统**，即使它是**早已存在的传统**。”

一百年前，著名的数学史学者大卫·尤金·史密斯和路易斯·查尔斯·卡宾斯基写道：

> 我们是如此熟悉那些带着引人误解的阿拉伯名称的数字，而它们在欧洲和美洲是如此广泛使用，我们很难意识到它们在商业交易中普遍被接受只是最近四个世纪的事，而今日有相当大一部分人还不知道这些数字。

史密斯和卡宾斯基继续指出，尽管其他所有系统是那么粗糙和笨拙，但这套系统却历经艰辛才成为全世界的标准系统，这是多么奇怪

的事。我们很容易忽略我们所有人都在使用的系统是如此近期的产物，但令人惊奇的是，这套系统传到欧洲已有数百年时间。

在斐波那契的时代，有许多关于计算的文本。然而，那些文本是为了想学数论（*arithmetica*）——关于数字和计算的理论及哲学——的人撰写的学术著作，或者是为了想精通教会历书的人所写。数学史家沃伦·凡·艾格蒙在博士研究中发现，那些书当中有许多并未在书名里使用abacus（计算）一词，而偏好用algorism（算法）这个词，这个词更直接地提示了关于印度-阿拉伯数字及其算法的说明。

尽管可能是斐波那契的著作将阿拉伯数字引入欧洲社会的某些地方，但也有可能意大利的旅行者和商人早已知道那些数字。而可以确定的是，在斐波那契的《计算书》之前差不多半世纪，有其他讨论阿拉伯数字的书籍。12世纪西班牙的《圣经》注解者、科学家及犹太拉比伊本·以斯拉撰写了《单位之书》来描述阿拉伯符号，并撰写《数字之书》描述有位值和零的十进制系统。他的著作对将阿拉伯数学的信息传递给商人大众没有太多贡献，但确实有助于让欧洲的学者注意到阿拉伯数学。大约在伊本·以斯拉撰写关于位值和零的著作时，著名的托雷多翻译学校（Toledo School of Translators）一位重要的翻译者希斯帕伦西斯，在他的著作《实用算术的计算书》中，写下被认为是已知西方最早对印度位值记法的描述。

20世纪早期的英国数学家及史学家W. W. 劳斯·鲍尔在其《简要数学史》（*Short Account of the History of Mathematics*）一书中告诉我们：

> 尽管是李奥纳多（斐波那契）将阿拉伯数字引进商业事务，但盛行于东方的阿拉伯数字知识，在旅行者和商人间可能已不罕见，这是因为基督徒和穆斯林之间的往来十分密切，足以使双方互相学习一些语言和惯例。我们也很难想象，意大利商人对他们最好的一些客户所使用的记账方法一无所知；我们也不要忘了，有很多在穆斯林那里当奴隶的基督徒，后来逃了出来或者被赎回。

14世纪伊始，佛罗伦萨的银行家被禁止使用阿拉伯数字，而16世纪之前，这类数字尚未通行。1299年，佛罗伦萨市议会颁布交易所法规，宣布在金融账目中使用印度数学系统为非法，要求所有金钱记录必须以文字记账，如同今日银行支票的规定。之所以采取这种做法是基于安全理由，以防止利用在数字上方或下方简单加一笔的方式，把0篡改成6或9。

交易所法规无法影响市场、市集和交易所中的日常交易，人们可以使用印度系统做计算，最后将账目转换成罗马数字。实际上这项法规对新兴起的印度系统的使用并未造成太多阻碍。更确切地说，是纸张和潦草计算所需的可擦拭材料的昂贵费用，阻碍了印度系统的发展。执行了长除法这类必须划掉很多步骤的计算之后，纸张无法用于后面的计算。而旧系统只需要一块计算板、算盘或沙板，不需额外的费用。

算术图

在图5-3中，我们发现毕达哥拉斯坐在计算板前，而波埃修斯在用印度数字做计算。为什么是毕达哥拉斯？因为在中世纪时，人们误以为毕达哥拉斯是算盘的发明者。


PVS A
RITH
TICAE

图5-3 算术图（*Typus Arithmeticae*）。格雷戈尔·赖希的著作《哲学珠玑》（*Margarita Philosophica*）中的木刻插图，该书最早问世于1503年，在高等学校中作为百科全书式教科书五十年之久。这张插图描绘了两个计算者（据称是毕达哥拉斯的人物使用计算板，波埃修斯以印度系统计算）在算术化身的女性人物见证下进行一场竞赛。资料来源：国会图书馆

哈里发的故事为许许多多逸事奇谈提供了背景，我们有时忘了那些多半是带有异国情调的古老文明的西方民间传说。或许这是因为巴格达的阿拉伯人从他们的征服地和波斯湾港口累积了难以置信的财富，在东方的中国、印度、俄罗斯与西方的整个欧洲之间推动了他们的贸易。下述关于印度数字如何传入阿拉伯国家的故事也许是虚构的，它出自13世纪中叶的著作《学者年表》［*Ta’rikh al-hukama*（*Chronology of the Scholars*）］，该书是伊本·奇夫提（Ibn al-Qifti）引述更早期的资料撰写而成。

哈里发曼苏尔（al-Mansur）在巴格达的皇家官邸接待了印度大使，那年是771年。大使送给哈里发的礼物是《婆罗门修正历书》，一本由印度数学家及天文学家婆罗门笈多在约一百五十年前以梵文写成的天文学著作。曼苏尔大力提倡文学和学术的传播，因此下令将《婆罗门修正历书》翻译成阿拉伯文。这个故事很可能只是传说，因为阿拉伯天文学应该有许多不同的来源。无论是否为传说，促使阿拉伯学者致力于天文学的可能就是这本书。

我们的零，作为一个数和一个占位符号，最早在书籍中出现可能是在近公元628年的某个时间。在婆罗门笈多的《婆罗门修正历书》里，我们第一次发现使用零和负数（负债）、正数（资产）的法则。婆罗门笈多将零标记成一个单一的黑点，用以表示从一个数减去本身之后所得的数。零不只是一个占位符号；或许这是有史以来第一次，有一个数代表nothing（无）。

我们对婆罗门笈多所知不多。他可能出生在印度南部的比拉马拉，但我们确知他在年少时期，移居至东南方四百公里处的乌贾因（Ujjain），在这个数学与天文学研究中心工作，那是6世纪的印度数学家及天文学家阿耶波多创立的。他研究高深的天文学，并发展了求平方根与解二次方程式的算术法则。

对于印度数学家阿耶波多生存的年代，我们同样了解不多，因为关于当时印度著述的历史记录相当稀少。古印度时期，印度人相信万

物皆有神性或灵之起源，包括科学也是。天文学和数学归因于梵天（Brahma），他创造了整个世界，因而印度文化会忽视真正做出科学发现的人类。

鲍尔认为阿拉伯人一旦离开沙漠定居在巴格达和大马士革等城市中，便开始染上他们缺乏免疫力的疾病。当时，希腊人和犹太人的医学远比阿拉伯人进步。就这点而言，以亚里士多德和盖伦（Galen）的研究为基础，他们对所有科学知识的研究也远比阿拉伯人进步。因此，哈里发鼓励希腊和犹太医生前来教授他们科学，并保存他们的技艺传统。“阿拉伯人的科学知识，”鲍尔说，“最早来自在巴格达照顾哈里发的希腊医生。”

约公元800年时，哈里发哈伦·拉希德下令将希腊著作翻译成阿拉伯文，他的后继者哈里发马蒙依循了这项指示，派出代表团到君士坦丁堡和印度馆藏丰富的图书馆，抄写了数百份希腊文和印度文著作。代表团返国后，一大批叙利亚书记员奉命将欧几里得、阿基米德、阿波罗尼斯和托勒密的著作，翻译成阿拉伯文和叙利亚文。

多亏他们收集了那些著作，因为它们是现存仅有的复本。鲍尔也提到，令人好奇的一点是，丢番图的著作在将外来作品翻译成阿拉伯文的传统刚兴起的前一百五十年间，似乎没有引起注意。那个时候，阿拉伯人已经相当熟悉自己的代数记法了。

---

[1] 15世纪之前，abacus这个词指的是机械工具之外更广泛意义的计算，它意指“做算术”。

第6章 阿拉伯数字的错误叫法

1983年，苏联发行了阿尔-花拉子密诞生一千两百周年纪念邮票，邮票上阿尔-花拉子密的肖像（参见图6-1）显示的是一个蓄胡子的男人，皱着眉，眼神迷离。我们可以知道某个9世纪的人长什么样子，但对他的生平却知之甚少，这不是很奇怪吗？事实上，我们几乎不知道他真正的模样。阿尔-花拉子密是他那个时代最伟大的阿拉伯数学家，他从婆罗门笈多《婆罗门修正历书》的阿拉伯文译本习得印度数字，并撰写了一本利用这些印度数字来做算术的教科书。[1]在约820年，阿尔-花拉子密撰写了《还原与对消计算概要》，这是一本关于如何解方程式的书（特别是解二次方程式的正根）。algebra（代数）一词，便来自这本书的阿拉伯文书名：*Hisab aljabr w'almuqabala*。12世纪中叶，这本书被翻译成拉丁文，书名为*Algebra et Almucabala*，这是algebra这个词具有今天的含义的由来。该书是现存最古老的阿拉伯代数著作。

图6-1 阿尔-花拉子密

阿尔-花拉子密以阿拉伯文撰述的原始算术书已不复存在。12世纪之初，该书出现在西班牙，由英格兰的翻译者切斯特的罗伯特（Robert of Chester）翻译为拉丁文。这部译本（发现于19世纪）和那段时期的其他著作，是已知最早将印度-阿拉伯数字传入欧洲的书籍，或许比斐波那契的《计算书》早了差不多一世纪。

阿尔-花拉子密的《印度数字计算》（*On the Calculation with Hindu Numerals*）成书于近公元825年的某个时间，9世纪时可能就是借由该书将印度的记数系统传遍阿拉伯国家，而后通过一系列拉丁文译本把印度计数系统传入12世纪的欧洲。

786年至809年间的波斯，拥有先进的科学和艺术财富。即将迈入9世纪时，伊拉克的阿拉伯帝国阿拔斯王朝（Abbasid）第五代哈里发哈伦·拉希德在巴格达建造了一座图书馆兼翻译中心，它在接下来五百年的伊斯兰黄金时代变成重要的学术中心，被称为“智慧宫”。希腊文、中文和其他语种的天文学、数学、农业、医学和哲学作品，在智慧宫里被翻译成阿拉伯文。阿尔-花拉子密在这里进行研究，对所有源自印度的文献深感兴趣，包括婆罗门笈多的《婆罗门修正历书》，这本书在图书馆的虫害中幸存下来。在解读那些神秘的记号和将该书翻译为阿拉伯文时，他有了重大发现：一种比笨拙的阿拉伯算法更简单的算术方法。

截至当时，在美索不达米亚各处活动的阿拉伯人一直用下列方式来做算术：手指计数、算盘、复杂的罗马数字系统、将数字记成文字的繁杂体系，甚至精心计算恒星位置时也是如此。

阿尔-花拉子密可能在印度文书写的《婆罗门修正历书》中发现了一种极佳的方法，仅需十个符号，就能简单表示任何计数数字（counting number，即自然数）。他或许已经知道巴比伦人的六十进制（以60为基底），从而了解表示十进制的方法。虽然十进制在中世纪早期阿拉伯国家的商业活动中并不盛行，他可能仍将其视为一套先进的系统，即使未吸引全世界注意，至少值得学术界认真关注。

他或许已经知道，那套系统中的奇特黑点表示“无”（nothing），也即“无”的量。任何阅读《婆罗门修正历书》的人都会对这点感到困惑。书中将标记负值的数视为负债，他可能从这种概

念中得到了启发。一种全新的对象之无限集合进入思维世界，这些对象将小于无的量——负数——符号化了。

有一个让人半信半疑的故事提到，阿尔-花拉子密到印度旅行，在那里接触到婆罗门笈多的数学手稿。然而，更可信的是，印度天文学家康卡（Kanka）在770年造访巴格达的智慧宫，他从印度带来许多手稿，其中包括《婆罗门修正历书》。这个故事比较合理，因为阿尔-花拉子密当时是智慧宫的学者，他在康卡造访后约五十年才撰写了其著作《印度数字计算》。主要就是凭借这本书，印度记数系统传遍阿拉伯国家和欧洲。

当时阿拉伯人没有自己的数字系统。阿拉伯世界的地理区域中说阿拉伯语和希腊语两种语言，他们所使用的系统，要么是希腊的字母系统，要么就是源于希腊形式而主要使用阿拉伯文字来表示数字的系统。那些新的数字有时被称为印度数字，有时则称为阿拉伯数字。斐波那契在他的《计算书》第一章开头，明确地将那九个数字称为印度数字。那段文字翻译如下：

> 九个印度数字为：
>
> 9，8，7，6，5，4，3，2，1。
>
> 利用这九个数字，以及阿拉伯人称为zephyr的记号0，任意数字都可以用下述方式写出来。一个数字是数个幺元（unit）的和，通过将数字相加，可以无止境地构造出新数字。第一，由幺元所组成的数包括一到十。第二，利用十可构造出十到一百这些数。第三，利用一百可构造出一百到一千这些数。第四，利用一千可构造出一千到一万这些数。因此，当我们不断地持续下去，任意数字都可以利用之前已造出的数构造出来。

用来表示这九个数字的书写形式差异很大，很可能因此造成混乱。然而，到了下一世纪，许多书写形式已经非常接近今日我们所用的标准符号。但阿拉伯的征服（Arab conquests， 623年至11世纪50年代）之后，阿拉伯天文表持续使用字母数字达数百年。在伊斯兰数字摘记中，没有统一使用印度-阿拉伯数字。

---

[1] 阿拉伯名字中的al（阿尔）意思是“来自出生地”。Khwarizmi（花拉子密）是今日乌兹别克的一个省。

第7章 一本文献引发的争论

9世纪时，印度数字仍然太新奇、太怪异，以至无法从修道院和学者的喧闹中广泛传播开来。毕竟欧洲人并不知晓任何一个包含零的数字系统，零这个符号可用于书写无限大范围的数，而同时也代表“无”。巴比伦系统没有零，希腊系统没有，罗马系统也没有。

这不是说商业活动和旅行者似乎未将这些数字带入欧洲。其实是有的。原因在于那个怪兽，也就是零——这个令人疑惑的陌生者，让这个新系统推迟了三百多年才被普遍接受。今日，我们迅速地接受各种创新产品，却很少注意到它们如何彻底地影响我们的生活——计算机芯片、手机、卫星定位系统、延长生命的医疗器材。难以置信的是，欧洲花了三百多年才理解一个有史以来简化人类生活最伟大的设计概念。三百年！中世纪盛期的伽利略、笛卡儿或牛顿在哪儿呢？

困难之处在于区别占位符号与数字。接受零是一个代表没有量存在的数字，这个想法无比大胆。举例来说，数字二相当容易理解。它代表“2之所以为2”（two-ness）或代表计数两个对象的数字。但“0之所以为0”（zero-ness）呢？一个没有对象可计数的数字？这可能会是什么意思呢？然而，以零作为占位符号的概念，错综复杂与将零视为一个代表“无”的数字的观念联系了起来。令人困惑的是，表示空位的符号，与表示**没有对象**可计数的数字，两者是相同的。缺少用来表示一个位置没有对象的第十个符号，印度九个符号的概念就行不通。这是巴比伦位值系统的部分难题。

然而，斐波那契面对的是在码头、市集和法院中的商人。他在其著作《计算书》中表明印度数字对那些商人来说是新奇的事物，因为他写道，他父亲（一位公证人）在他小时候带他到贝贾亚（在今日的阿尔及利亚），他发现了这个系统，并在那里学会计算之术。他写道：“拉丁民族”不懂印度的算术方法，而其他常见的计算技术，如算法和尖点体，都是“相较于印度方法而言错误的方法”。他真诚地相信这些看法。

身为同时代最伟大数学家之一的斐波那契，对于明确论及印度人以那九个数字来计算的技术的早期著作显得一无所知，这点似乎令人讶异。难道他不知道阿尔-花拉子密的《印度人的计算》（*On the*

Calculation of the Indians）吗？该书在前一个世纪早期就译成拉丁文了。他不知道976年西班牙阿贝达圣马丁修道院的手稿吗？那份手稿中提到：“我们应当知道，印度人具有最难以捉摸的天赋，让其他民族的算术和几何学相形失色……”他不知道意大利北部的托雷多翻译学校翻译了手稿？他不知道根据九个印度数字而发明的吉尔伯特算盘？他怎么会不知道撰写于托斯卡纳的欧几里得《几何原本》有希腊-拉丁文译本，其中使用了东方形式的印度数字符号？距斐波那契的故乡比萨仅一百多公里的公证人，已开始使用印度数字。从托雷多到里昂到慕尼黑再到爱尔兰，与计算有关的拉丁文书籍，都提到了运用九个字母的印度数字记数法，并说明了如何用印度数字来表示所有的数。

然而，究竟是谁将运用印度数字的计算方法引进欧洲并产生了广泛影响，答案并非显而易见，证据比较庞杂。斐波那契在比萨受教成为商人，在学校里，他学习用罗马数字在计算板上记录和计算。作为学徒时，他学会如何计算货品的价格，如何称重和量尺寸，以及如何转换等值的金钱。等他到了贝贾亚，已经能用计算板处理例行的商业算术。学会了印度数字和相关算术运算的方法之后，他发现印度的计算法比在比萨所用的那些方法好。当他回到比萨后，就不用再研究印度系统的拉丁文本了，他在贝贾亚已经学会了那套系统。这或许可以解释，为什么斐波那契没有提过任何关于利用九个数字来计算的印度技术的早期著作。

直到近期，中世纪研究者始终确信斐波那契的《计算书》是将现代算术引进西方的启发之作。2004年，数学史家拉法埃拉·法兰契推许此书为“意大利计算教学最重要的文献”。数年前，另一位著名的史学家伊丽莎贝塔·乌丽薇宣称，以托斯卡纳方言写成的计算著作，主要取材自斐波那契的两部著作《计算书》和《实用几何学》。而回溯至1980年，凡艾格蒙编目整理了到14世纪中叶为止，直接源自《计算书》的大量计算著作，证实了印度数字在意大利的传播及这些数字与斐波那契的关联。

接着（1989年之前）是吉诺·阿里吉的发现：他在佛罗伦萨的瑞卡迪纳图书馆找到一本以翁布里亚（Umbria）方言写成的著作《计算之书》（*Livero de l’abbecho*）。[1]该书无疑是1289年（或前后一年）在翁布里亚成书的，作者不详。这是现存最早以方言写成的计算著作，可能是仿效一部更早期的版本撰写的，它开宗明义地写道：

> *Questo ène lo livero di l'abbecho secondo la oppenione de maiestro Leonardo de la chasa degli figluole Bonaçie da Pisa.*
>
> 这是一部计算书，附议比萨的波那契（Bonacci）家族的李奥纳多大师之见。

从这段文字以及其他可信的证据，法兰契提出，无论这位翁布里亚大师是谁，这位作者可能变更了写法以适应其读者的需求。斐波那契可能就是翁布里亚大师本人，而该书或许其实是佚失的《简短之书》［*Liber minoris guise（Book in a smaller sense）*］，斐波那契在其《计算书》中引用了这本书，因此我们知道那也是他的著作。果真如此，他是西方算术之父的看法将确凿无疑。但数学史家贺鲁伯认为，如果我们审慎地读完序言，会“发现里面有些内容绝非出自斐波那契之笔”。法兰契认为，仅凭《计算之书》刚开始的部分并非来自《计算书》这一点，不能说它看起来不像《计算书》。

由于敬重贺鲁伯的苦心阅读，法兰契改变了原先对斐波那契的实际贡献的部分看法，主张那些计算著作的“作者可能曾接触与李奥纳多所用的那些文献不同的阿拉伯文献”。她目前正研究两本在13世纪末或14世纪初写于比萨的计算论著，两书深受《计算书》前八章的启发。[2]

印度数字传入西方无疑是从10世纪末以后开始的，这不必然意味着印度人的计算方法早在斐波那契之前就已传入。不过话说回来，根据另一位著名数学史家查尔斯·伯内特的看法，许多12世纪的印度计算著作指出斐波那契并不是先行者。

斐波那契在《计算书》的序言中告诉我们，他是在与父亲旅行时，在埃及、叙利亚、希腊和普罗旺斯遇见一些商人，学到了贸易使用的那九个印度数字。普罗旺斯？普罗旺斯不是在西欧吗？与普罗旺斯有贸易往来，却未启发意大利的计算算术，这怎么可能？

贺鲁伯谴责所谓的“巨著的原则”，那个原则声明每一本书要么是原创，要么就必须将它的见解归于某部已不复存在的名作。他写道：

> 《计算书》的某些段落显示，计算数学的发端，必定可追溯到斐波那契时代已经存在的环境——一个他熟知的环境，在那个环境里，他可算是早期卓越的倡导者，但并非创始者。

斐波那契对尚不为人所知的印度数字并无任何贡献。然而，他是新颖且困难的概念的杰出阐述者。他作为阐述者的天赋，或许影响了这个新系统从意大利扩散到欧洲其他地区的过程。到了13世纪中叶，已有多部拉丁文著作将这个新系统引进北欧。举例来说，《算法之歌》是一本广受欢迎的专论，法国方济会修士维尔迪厄在1240年撰写了这部著作，书中以两百四十四句扬抑抑格六音步诗说明了计算的方法：

> 算法是这样的。
>
> 这个称为算法的新技术，其中
>
> 利用五个数字的双倍
>
> 0 9 8 7 6 5 4 3 2 1
>
> 我们受惠于印度人的发明。

印度数字在12世纪和13世纪的学者群中广受欢迎，因为在修道院的手稿里经常可见这些数字。托雷多翻译学校的塞维亚的约翰（John of Seville）所翻译的一部拉丁文译本，在切斯特的罗伯特的译本之后随即问世，接着在1143年，切斯特的罗伯特的译本删节版收录进德国南部萨勒姆修道院（Salem Abbey）的图书馆，这是阿尔-花拉子密的《代数》传入欧洲北部最早的证据。[3]还有在巴黎的新大学任教的萨克罗博斯科，他在1240年撰写了《算法》一书，这本教科书讨论印度数字及如何用这些数字来计算，全欧洲都广泛使用这本书（参见图7-1）。

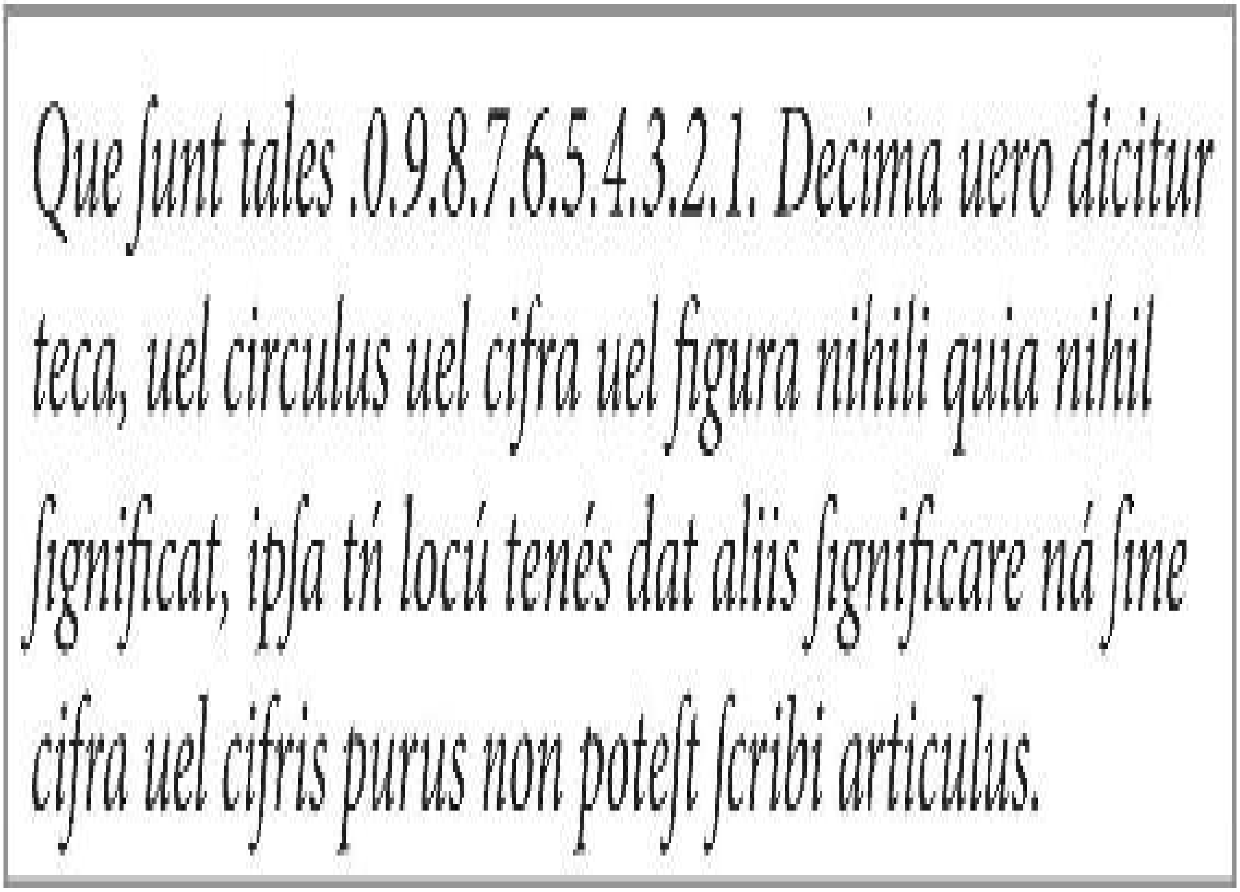

Que ſunt tales .0.9.8.7.6.5.4.3.2.1. Decima uero dicitur teca, uel circulus uel cifra uel figura nihili quia nihil ſignificat, ipſa tñ locú tenés dat aliïs ſignificare ná ſine cifra uel cifris purus non poteſt ſcribi articulus.

图7-1 摘自萨克罗博斯科《算法》1523年版的一段。第八行以下翻译如下："要知道，有九个数字符号对应九个幺元，如下所示：0，9，8，7，6，5，4，3，2，1。第十个称为theca或*circulus*或cifra或*figura nihili*，因为它代表无。但当它被放置在适当的位置时，会赋予他者数值。"资料来源：托马什计算历史图书馆

因此，看起来似乎在斐波那契撰写《计算书》之前两三个世纪，关于这些新数字的消息已传遍了欧洲。这是新鲜事物，但非实际算法，因此，没有多少人使用。一个可能的原因是它被误解了。有人尝试让罗马数字顺应到位值系统中。罗马字母经常用于位值系统，不考虑零的概念。例如，数字16在罗马系统中写成XVI；就这里的位值意义而言，是表示XVI与XIV并不相同。别忘了，罗马人也使用计算板，上面区别了代表十的列、代表五的列及代表幺元的列。

当商人和会计员应该已经明确看到印度系统应用在计算板和算盘的位值上，并且每天在市场中以位值模式的概念谈论数字，中世纪的欧洲人怎么会无法理解这套系统的价值？

是什么原因让他们未能认识到这套印度系统的好处？一个可能的答案是，它或许比我们想象的更让人畏惧。想象一下，在我们已经如此熟悉印度系统之后，要学习使用比如希伯来数系这样的系统，会多么费力。唯有当你有机会使用一个新系统，你才能了解它的好处。

到了11世纪末，关于印度系统的信息借由在吉尔伯特算盘的算子上标记印度数字，传遍了欧洲。所以为什么我们如此赞同斐波那契将印度数字引进欧洲这件事？

在史密斯和卡宾斯基于1911年出版《印度-阿拉伯数字》（*The Hindu-Arabic Numerals*）之前，关于现代数字起源的争论已风起云涌了近百年。史密斯和卡宾斯基表示，“在商业交易里普遍接受［印度数字］只是最近四个世纪的事”。韦氏字典列出0至9为阿拉伯数字，但我们发现一份662年的手稿残篇中提到这些数字的印度根源，这份手稿出自住在幼发拉底河肯尼锡（Kenneshre）的修道会的叙利亚学者，也就是尼西布斯（Nisibus）的塞维鲁主教之手。这份残篇藏于法国国家图书馆（MS Syriac [BNF]， No. 346），是现存已知来自印度之外最早的印度数字参考文献：

> 我将省略关于印度人的科学的所有讨论……他们精妙的天文学发现，比希腊人和巴比伦人的那些发现更具独创性的发现，以及难以描述的他们可贵的计算方法。我只想说这种计算是借助九个记号来完成的。如果那些因为本身讲希腊语，就认为自己已经达到科学极限的人会读印度著作，他们将确信，即使时间上晚了一点，仍有其他人像他们一样了解某些重要的事物。

所以，塞维鲁主教确已知道这九个记号，他是7世纪时在叙利亚的希腊哲学和科学的主要传播者之一。上述译文残篇除了挖苦评论其他讲希腊语的人，还宣称利用九个记号来表示所有数字的想法来自印度人。

塞维鲁认为，那个系统是从印度往西传遍了波斯世界。更近期（若我们可以说1977年是最近的话），研究中世纪占星学的史学家理查德·李梅（Richard Lemay）认为，连同《还原与对消计算概要》在内，阿尔-花拉子密的《算术》在12世纪间就有三种不同的拉丁文译本。他写道：“阿尔-花拉子密的《天文表》（*Astronomical Tables*）是最值得注意的一个途径，经由这个途径，西方得以知悉这个印度-阿拉伯数字系统。”那九个记号的原创者是印度人这点，极可能是出自阿尔-花拉子密的《算术》。

在阿拉伯国家，这九个数字也称为“印度字母”或“数字”。少数确凿的一份文献是10世纪《黄金草原和珠玑宝藏》的记述，这部于957年问世的三十卷著作是冒险家兼说故事者马苏第毕生的心血。马苏第在第一章里写道，他选用这个书名是“为了读完内容后激发求知欲和好奇心，并且让心灵渴望熟悉历史”。马苏第是个富有好奇心且热衷于探究的人，他收集了波斯人、印度人、犹太人和罗马人的历史故事，以及东方文明的文化。他出生于巴格达，曾造访印度、锡兰（今斯里兰卡），横越印度洋到马达加斯加，上行红海回到埃及、巴勒斯坦和叙利亚。我们发现他在926年到过以色列太巴列（Tiberias）的加利利海（Sea of Galilee）附近；到了943年，出现在地中海周边的安提阿（Antioch）或基利家（Cilicia）附近；两年后，现身大马士革。他将其著作视为献给“国王和学者的一份礼物。内容论述了每一个可能有助于学习或引发学习好奇心的主题，并论及所有在时间推移过程中出现过的知识”。马苏第在其著作中通篇使用印度-阿拉伯数字，该书一开始是一段关于天文学家侯赛因（Hosaïn）的描述，他汇编了天文表，并讲述关于地球周长和直径的错得离谱的论据。

阿贝达伊雷瓜（Albelda de Iregua）是西班牙北部的一座小城镇，阿贝达圣马丁本笃会修道院的废墟矗立此地。在其10世纪全盛期，它是西欧最重要、最先进的文化中心，很可能是因为它就在埃布罗河（Ebro）沿岸的贸易路线上，连接西北部的卡斯提尔（Castile）与地中海。阿贝达圣马丁本笃会修道院有一座馆藏丰富的图书馆，收集了西方可得最丰富的中世纪西班牙文献，包括西欧最早关于1到9的阿拉伯数字的记录。塞维亚的伊西多尔（Isidore of Seville）于976年在该修道院撰写的拉丁文手稿《词源》（*Etymologiae*），已经约略显示出除了4之外那些数字的现代形式。经过多年的演化，基本形式和特征趋于标准化。要精准指出趋于统一的正确时间点或许无法做到，不过我直觉认为这要归因于当时逃难的数学家。

现存最早关于印度算术的阿拉伯著作是阿布哈桑·阿尔乌几里德的《印度算术原理》（*Kitab al-fusul fi’l-hisab al-hindi*），约952年时撰写于大马士革。使用印度数字最早的阿拉伯实例是两份写在莎草纸上的鳄城（Crocodilopolis）法律文件，该城（埃及最古老的城市）是由目睹当地居民礼拜活鳄鱼的希腊探险家所命名的。这两份文件以阿拉伯数字标注日期：873—874年和888—889年；接下来最早的一些例子要到11世纪才出现。到了12世纪，如摩洛哥数学家伊本·亚萨敏的描述，西边的伊斯兰世界与东边的伊斯兰世界书写印度-阿拉伯数字的方式有了明确的差异。已知最早以意大利方言写成的使用印

度-阿拉伯数字的著作是《计算新书》（*Libro di nuovi conti*），约成书于1260年，但现已不存。

根据所有文献资料和证据，我们能对现代数字的起源做出什么结论？它们源自印度？阿拉伯？中国？法兰西？专家对印度-阿拉伯数字的起源争论不休近两个世纪。其中一位专家是法国数学家兼史学家沙勒（Michel Chasles），他满怀爱国心，为印度-阿拉伯数字起源于法国这一个荒谬案例争辩，而其所依据的文件显然是伪造的。

---

[1] 我不知道发现这本书的确切时间，但必然是1989年之前。一些广受欢迎的著作认为发现者是法兰契。然而，在我与法兰契的谈话中，她说那本书实际上是阿里吉发现的。

[2] 感谢法兰契阐明了斐波那契对印度数字在意大利传播所产生的影响的争论。

[3] 12世纪时，萨勒姆修道院图书馆是欧洲最重要的图书馆之一。

第8章 符号起源地的众说纷纭

弗兰-卢卡斯（Denis Vrain-Lucas）因无法合法收集到想要的文件而苦恼，所以在巴黎多座图书馆割下旧书扉页偷取古董纸张来伪造文件。他使用特殊的自制墨水，仔细模仿不同的笔迹，然后把赝品（信件和文件）卖给未起疑的手稿收藏家。

他是书记官，也是业余史学家，他充满热情地收集富有重大历史意义的手稿。1855年左右开始，超过十六年的时间，弗兰-卢卡斯卖出两万七千多件伪造亲笔签名的赝品，其中许多是卖给他最喜欢骗的目标沙勒，后者自1861年开始在九年间支付了数十万法郎。附有帕斯卡尔、伽利略、笛卡儿、牛顿、拉伯雷和路易十四亲笔签名的信件很可能被认定是真迹，而弗兰-卢卡斯在手稿收藏界累积了相当高的声望，让他能用那些荒谬可笑的东西冒充真品。

天真的沙勒购买了克丽奥佩特拉写给马克·安东尼的多封签名信（居然用的法文！），以及亚历山大大帝的一封签名信（也是法文！），而帕斯卡尔、牛顿、伽利略之间的来往书信也全是用法文，其中证明是帕斯卡尔发现了万有引力定律。牛顿在《原理》一书中关于万有引力的描述是在帕斯卡尔死后二十五年才发表的，所以任何这类有帕斯卡尔签名的信件当然让人吃惊。然而，1867年，沙勒带着他珍藏的信件来到法国科学院提出证据，尽管某些科学院成员心生怀疑，但其他人则沉溺于民族自尊心而假定那些东西是真的。1869年，弗兰-卢卡斯因伪造罪受审，被判两年监禁，但未强制他须赔偿沙勒的损失。不过，即使证据确凿，显示那些手稿是骗人的，沙勒仍坚信它们是真品。

有一位科学院成员始终坚定不移地怀疑这件事。这个人的全名是古列尔莫·利布里·卡鲁奇·达拉·索马埃布尔伯爵（Count Guglielmo Libri Carucci dalla Sommaja）。19世纪40年代的大部分时间里，沙勒和利布里在法国科学院的会议中针锋相对，争论的焦点多半是数字起源于哪个国家及代数的起源。

沙勒认为，在5世纪之前，法国已经有一种用于计算的十进制位值系统，波埃修斯的《算术》证明了这点，该书似乎使用了带有阿拉伯数字的乘法表。后来的学者强烈怀疑这部使用了阿拉伯数字的著作不

是原始版本，但沙勒认为斐波那契的《计算书》是受到阿拉伯作者的影响。

与此同时，沙勒的对手利布里伯爵刚出版了《意大利科学史》中的一卷，其中论述了阿拉伯作者所使用的算术和位值记法起源于印度的问题。沙勒质疑利布里的观点，后者认为他们的现代数系是借由意大利斐波那契的著作传到欧洲的，而前者则认为是法国人韦达将现代数系传到了欧洲。这成为一场公开的论战，打着强烈对立的社会政治旗帜针锋相对。

利布里在三十八岁时被任命为法国图书馆的总督察，这激发了他处理珍本著作的长久兴趣，也使他难以克制冲动偷走了大量珍稀手稿。到了他四十五岁时，政府发出了对他的逮捕令。他带着两万多部珍稀书籍和手稿逃往伦敦，其中有一些书是他年轻时从佛罗伦萨的劳伦佐图书馆偷来的。任职于学术机构的杰出人士出现这种怪异的偷书行为可能令人费解，但这种行为在19世纪的法国并不稀奇。

19世纪大部分时间，十进制位值记法起源于印度的说法一直受到质疑。接着，1907年，在英属印度夏都西姆拉的印度政府教育部门任职的英国业余印度学研究者，在《孟加拉国亚洲学会期刊》发表了一篇文章。在那篇文章中，乔治·鲁斯比·凯伊声称数字和十进制位值系统不可能起源于印度。他的论点部分是因为对巴赫沙里手稿的错误判断及对这份手稿的年代判定。1881年，一个农夫在巴赫沙里村（位于今巴基斯坦）附近，挖掘出这份以梵文和古印度方言普拉克利特语（Prakrit）书写在桦树皮上的文献。这份手稿被发现时是断简残篇，原来可能有数百片桦树皮，由于处理不慎而腐烂，最后仅余七十片。

毋庸置疑，巴赫沙里手稿的记述跟我们的现代数字和位值系统存有差异。然而，这份手稿的年代始终具有争议。有些学者估计是公元400年，有些人推定是公元700年。凯伊声称可能的年代是13世纪初，但他在其1907年影响深远的文章中写道：“我们可以进一步如实指出，在印度数学的整个发展过程中，毫无迹象显示在10世纪之前曾运用任何位值的概念。”他暗指争论中的记法是起源于阿拉伯国家。要么他不了解这种位值记法，要么可能是他为了保护英属印度的殖民利益而排除了位值记法起源于印度的可能性。近期的学术研究推断，巴赫沙里手稿的年代介于公元前200年至公元300年间，他们的理论依据是这份手稿所用的语言自3世纪后已不存在。

巴黎第七大学法国国家科学研究中心（CNRS）数学史家阿加莎·凯勒如此评价凯伊：

> 我们在这里遭遇到科学史中一个奇怪而又熟悉的情形，这种事常常发生在科学故事中，这就是：有人拒绝接受事实。我们如何理解凯伊的态度？他无疑接触过证实这样的说法不可信的文本。

阿耶波多知道十进制，婆罗门笈多也知道。《昆耶舍论疏》（*Vyasa-Bhasya*）是五六世纪间昆耶舍以梵文书写的瑜伽论著，该书中有一个数学模拟的例证：“同样的数字‘1’，放在一百的位置代表一百，放在十的位置代表十，而放在幺元的位置代表一。”因此，远在阿拉伯人之前，印度人就了解十进制。中国人同样知道十进制。所以，凯伊如何能否认我们的数系起源于印度？

在后续的论文中，凯伊写道，由于有这么多假造时间的数据，印度-阿拉伯数字表示法的历史错综复杂。他指出西方以前对数学的了解，主要归功于早期的印度数学家。凯勒认为，就“西方知识”来说，凯伊的意思是希腊-拉丁的智慧，是由阿拉伯学者传到西方的。凯伊似乎深受位值起源于早期梵文著作的想法困扰。

若非凯伊的论文如此受到重视，一切都没问题。他的文章广受印度学研究者欢迎，权威数学史家也引用他的看法，甚至包括20世纪早期极为杰出的学者大卫·尤金·史密斯、路易斯·查尔斯·卡宾斯基、弗洛里安·卡裘利和乔治·萨顿。迟至1927年，著名的印度数学史家达塔仍写道：

> 这篇发表的论文的意义和重要性，对所有的科学史爱好者来说显而易见。而他们无疑因凯伊先生在解说和校订巴赫沙里手稿过程中所做的极大努力，心存感激。

我们使用的十个数字，包括代表一个空位的零，是源自印度并由阿拉伯人传到我们的世界的，这一点看来没有疑问。所以接下来本书会称这些数字为印度数字。

阿尔-花拉子密将印度数字描述为梵文符号，但他的论著《印度人的计算》在13世纪前尚未译为拉丁文传入欧洲，当时的商人仍用罗马数字来进行日常计算。这可能是我们的数字起源于何处众说纷纭的原因。印度人无疑的确造访过比叙利亚更西边的地区；470年，婆罗米人

受罗马宫廷之邀造访亚历山大城。然而，在那么早的时代，数字尚未被视为智识宝藏或科学珍宝，“反而像是在海港和港口城市流传开来的外族人的数字”。

无论真相如何，很可能在5世纪，印度数字已经由途经叙利亚的贸易路线传入亚历山大城。从亚历山大城这个连接欧洲的重要城市，数字往西方推进。

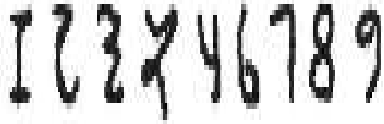
起初，无论是什么时候，这九个数字呈现混杂的形式；然而，到了13世纪初，斐波那契撰述《计算书》之后，那些数字的形式开始逐渐固定为我们今日所见的形状，只有少数例外。对于当时费力用罗马数字来记数和计算的欧洲人来说，这是天赐的礼物：认识到只需要十个符号，就足以表示在概念上无限的所有数字之集合体中的任意数字。罗马数字系统无法做到这点，因为对每一个十的乘幂，它都需要一个新的符号。

当然，无论他们使用哪一种系统，有时间又有毅力的人总是能进行计算。而他们向来有时间又有毅力！在这套有位值、零和简单算术的九个数字的杰出系统出现之前很久，东方的商人、天文学家和数学家已经以算术方法作为简单的工具，来进行艰难的计算。在差不多五千年间，各种形式的算术方法成为合宜有效的计算工具。10世纪时，算术方法往西方传播。

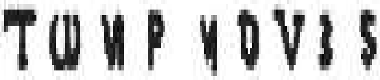
算术兴起于市场，而后发展到处理天文学问题。数字，商人的语言，必定来自用来描述数字的文字：一、十、百……在它们被各种记法符号化之前，这些都只是文字而已。一种描述大数字的简易方式，伴随着零的发明而出现，这时我们可以说“一零”“二零”“一零零”等等。一个单一的数字可以不断使用来表示无限多不同的数字。由此，我们了解到，利用我们已有的数串，可能写出无限多个数字。

虽然印度数字的信息通过商人和贸易人群广泛传播，但这些数字迟迟未成为标准形式。“一直要到16世纪，这些新数字才在学校和商业交易中彻底赢得胜利。即使迟至1543年，也就是哥白尼逝世那年，在他那年出版的著名著作《天体运行论》（*De Revolutionibus Orbium Coelestium*）当中，我们仍发现罗马数字、印度数字混用，甚至完全以文字书写数字的奇异现象。”

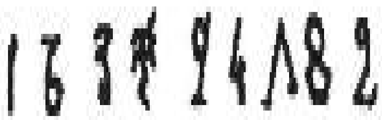
关于我们的数系起源的严谨研究已经非常多，但即使经过上百年的广泛学术探讨，我们仍仅能概略猜测数系的发端和演化（参见图8-1）。



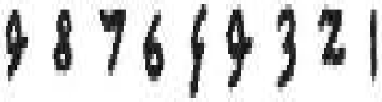
维希拉努斯抄本，写于西班牙阿贝达修道院，976年，欧洲使用印度–阿拉伯数字最古老的真迹手稿



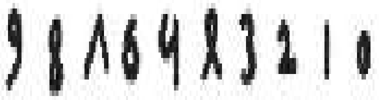
圣高尔手稿（St. Gall manuscript），10 世纪，苏黎世大学图书馆



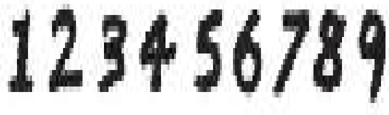
梵蒂冈 MS 3101, 1077



《计算书》，佛罗伦萨，1202年，国立中央图书馆，Maghabech C. 1, 2616, fol. 1v

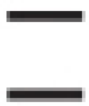
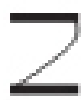


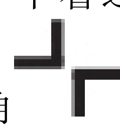
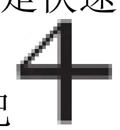
《算术》（*Psephephoria*），约1300年，梵蒂冈，gr.184, s XIII



《计算技术》（*De Arte Supputandi*），1522年，卡斯伯特·谈思陶尔（Cuthburt Tonstall）

图8-1 印度数字的形态演变。关于东西方数字的差异和形态变化更包罗广泛的记述，参见Charles Burnett，“Indian Numerals in the Meditarranean Basin in the Twelfth Century，with Special Reference to the ‘Eastern Forms’”，*China to Paris：2000 Years’ Transmission of Mathematical Ideas*，ed. Yvonne Dold-Samplonius，Joseph Dauben，Menso Folkerts，and Benno Van Dalen (Wiesbaden：Franz Steiner Verlag，2002)，237－284

我们可以推测数字的形态演变过程。在几乎每一种文化中，数值较小的数字的书写都始于点或线，这很可能是因为刚开始手边可取得的工具的关系——刀、凿子、细枝或芦苇。在木头、石头或黏土上书写速度很慢，而用墨水写在莎草纸、羊皮纸或纸张上，不需要高举刷子或笔，所以速度加快。从1到9的现代数字的每一个，就像有意安排的一样，都是可以用笔写出的单一标记。我们所用的“2”的形态自然演变过程似乎是，，2。虽然没有确切的证据，但这个猜测似乎是正确的。那条对角线看起来是无意间把墨水从顶部线段拖曳到底部线段所形成的，以免浪费时间把笔或刷子抬高到离羊皮纸够高的地方。同样，“3”可能是快速写下三条水平线段所自然形成的。有时那些线段写成直式，如果是这样，那些数字看起来会很像我们现代的数字，只是转了九十度。

但是“4”呢？它是怎么来的？乍看之下，它只有三笔——垂直线、水平线和斜线。把它想成两个角——也就是快速写下的四条短线——它最后变成一个未中断的单一标记。那条斜线是从垂直线拖曳到水平线而无意间形成的。

奇怪的是，现代数字的笔画数与那个基数本身没有直接的关联。早期数字符号的形态变化已无法追溯。举例来说（参见图8-1），从10世纪至16世纪，数字5看起来很像各种不同方向的h，有时倒栽葱，有时则不然。16世纪之前，数字4不像我们现代的4。

历史进程不是预期的结果，往往由偶然与巧合推动，无法预测，也不易控制。阳光温暖了浅水池里的黏质物，慢慢创造了生物生存的

条件，使地球上产生了生物。而地震埋葬了文明。非预期的驱动力改变了国家，这些驱动力可能是无法预知国家领导者是否英明，也可能是无法预测领导者的决定带来的是福是祸。若是签订第一次大战条约时能够更加理解它所造成的后果，第二次世界大战也许不会发生。换句话说，如果希特勒小时候死于风湿热，整个20世纪可能出现非常、非常不同的结果。理性的过程扮演了某种角色，但只是串联起偶然性与后果交织的错综复杂情况。一个偶然性使得某个人出生，而另一个人的生命太早结束；一个偶然性使得自然剧变摧毁了解答一项关键问题的线索，而另一个偶然性使得自然剧变创造了一条线索；一个偶然性让某份文件佚失，而另一份文件寻获。人类命运时间轴上的趋向，时而枯燥乏味，时而精彩绝伦，而且看起来就像外海上的天气一样变幻莫测。

要处理古代文献中的记法是有困难的：无论多么认真地检验文献，对于记法是如何出现的这件事，总是会有某种程度的猜测——抄写员是否加入了手稿原来没有的什么东西？印刷业者是否用冷式排版（cold type）[1]来替代记法，以便更容易留间隔？作者思考的全过程，或是他们的作品如何影响他人，就是史学家找到真相的最好机会。有些时候，大多数历史专家并不同意彼此的看法。人类知识的发展，就像已故科学家的传记，涉及如此众多相互交织的科学、经济学、神学、政治的因果式的关系，而推测往往是将它们联系起来的最佳方式。毕竟没有推特可以告诉我们，数学的早期贡献者到底是怎么想的。

总有奇怪的事情发生，也总有好的事情发生。这就是历史运行的方式。

这个世界最卓越的数学史家未必同意彼此的观点。这很好，这使得问题可以进一步研究，而这不就是历史最令人兴奋的地方吗？埋在沉积物中数十万年的一个有机体，因地震而重见天日。修道院图书馆里一份千年的羊皮纸文献，被发现是散佚已久的数学著作。一部因熔岩崩塌而保存了数百年的手稿出土，诉说一个真实的故事。历史因始料未及的惊奇而获得修正。这些事在每一个世纪上演着。

---

[1] 使用机械式铅字打字（打字机）、照相打字、计算机照相排版等不须经过热力处理，即可完成检排作业的排版方式。——编者注

第二部分 思维演化的历史

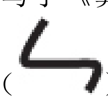
时间再回到印度数字传入欧洲之前。

主要的创始者

这些创始者多半是最早在著作中使用符号，或最为知名者：

丢番图（Diophantus，205±15—290±15年）

亚历山大城的希腊人，数学家。

约250年撰写了《算术》（*Arithmetica*）一书。最早以符号来代表减（）和未知数（）。

希帕蒂亚 （Hypatia，约350—370年）

希腊人，数学家。

首位杰出女性数学家，《算术》一书注释者。

阿耶波多 （Aryabhatta，476—550年）

印度人，数学家、天文学家。

使用字母来代表未知数。

婆罗门笈多 （Brahmagupta，598—668年）

印度人，数学家、天文学家。

可能是最早把零（一个小黑点）当作数字的作者（621年）。撰写了《婆罗门修正历书》（628年），书中以缩写来代表平方和平方根，以及出现在特殊形式问题里数个未知数中的每一个。提出负数与正数的乘法法则。

阿尔-花拉子密 （al-Khwārizmī，约780—约850年）

波斯人，数学家、天文学家、地理学家。

智慧宫学者。著有《还原与对消计算概要》（简称《代数》，830年）。以各种不同的形式，用文字表述代数式。

比萨的达地师傅 （Maestro Dardi di Pisa，Jacopo）

意大利人，数学家。

1344年的未出版手稿*Aliabraa argibra*是最早以意大利方言写成的代数专论。

帕乔利 （Fra Luca Bartolomeo de Pacioli，1446或1447—1517年）

意大利人，数学家。

他的代数专著是最早印行的相关书籍，以阿拉伯文命名：《还原与比较》[*Alghebra e Almucabala*（*Restitution and Comparison*， or *Opposition and Comparison* or *Resolution and Equation*）， 1478年]。

丘凯 （Nicolas Chuquet，1455—1488年）

法国人，数学家。

著有《数学三篇》（*Triparty en la Science des Nombres*， 约1484年）。将代表幂次的各个记号标记为℞，℞^{2} 2……将平方根标记为℞。

威德曼 （Johannes Widmann，1460—1498年）

德国人，数学家。

在其1489年的著作《各种职业中快速且工整的计算》（*Behende und hübsche Rechenung auff allen Kauffmanschafft*）中，引进“＋”作为代表加的符号。

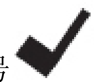
施蒂费尔 （Michael Stifel 或 Stefleius，1487—1567年）

德国人，数学家。

1553年出版《求根术》（*Die Coss*）。使用字母M和D分别代表乘和除。所以3②*D sec*①*M ter*②是表示 $\frac{3x^2z^2}{y}$ ，其中*sec*和*ter*代表第二个和第三个未知数。

鲁道夫（Christoff Rudolff，1499—1545年）

德国人，教科书作者。

著有《求根术》（*Die Coss*，1525年）。纳入符号，和来分别代表平方根、立方根和四次方根。

卡丹诺（Gerolamo Cardano，1501—1576年）

意大利人，医师、数学家、占星家。

1545年撰写《大术》（*Ars Magna*），解三次方程式和四次方程式。认可虚数解和复数解。

雷科德（Robert Recorde，约1512—1558年）

韦尔斯人，医师、数学家。

著有《砺智石》（*Whetstone of Witte*，1557年），广泛流传，将等号（＝）引进欧洲北部国家。

邦贝利（Rafael Bombelli，1526—1572年）

意大利人，数学家。

涉及求解三次方程式和四次方程式（1572年）。使用 $\underset{\smile}{0}$，$\underset{\smile}{1}$，$\underset{\smile}{2}$……来表示未知数及其平方、立方等。

克胥兰德（Guilielmus Xylander，Wilhelm Holzmann，1532—1576年）

德国人，学者。

古典学者，欧几里得《几何原本》和丢番图《算术》的拉丁文译者。

韦达 （François VIÈTE，1540—1603年）

法国人，数学家。

使用字母将数字表示为一般的数学对象，并应用相同的代数推理和法则。

西蒙·史蒂文 （Simon Stevin，1548—1620年）

法兰德斯人，数学家、工程师。

在他的《算术》（*L' Arithmetique*， 1585年）中，使用所谓指数计划（Index Plan）来书写指数，例如，x^2-3x+2会写成1②−3①+2⓪。

哈里奥特 （Thomas Harriot，1560—1621年）

英格兰人，天文学家、数学家、民族志学者。

令多项式等于零，从而发现：若a是小于五次的多项方程式的一个根，则$x-a$为该多项式的一个因式。

奥特雷德 （William Oughtred，1574—1660年）

英格兰人，数学家。

著有《数学之钥》（*Clavis mathematicae*， 1631年）。创造出超过一百个符号，但不到十二个沿用至17世纪之后。使用“×”来表示乘，冒号“：”代表除。

埃里冈 （Pierre Hérigone，1580—1643年）

法国人，数学家、天文学家。

著有《数学教程》（*Cursus mathematicus*， 1634年）。撰写了一部有六卷的代数教科书，几乎完全以符号写成。发明了“⊥”（垂直于）和“∠”（角）。

巴歇 （Claude Gaspard Bachet，1581—1638年）

法国人，数学家、语言学家、学者。

最早将丢番图的《算术》从希腊文翻译为拉丁文（1621年）。

笛卡儿 （René Descartes，1596—1650年）

法国人，数学家、哲学家。

著有《几何学》（*La Géométrie*， 1637年）。使用数字上标来标记多项式的正整数指数。将一个个乘方以数字等级排列。创立规约：保留字母表前面的字母来表示固定的已知量，后面的字母来表示变量或未知数。

沃利斯 （John Wallis，1616—1703年）

英格兰人，数学家。

著有《普遍数理》（*Mathesis Universalis*）和《无限算术》（*Arithmetica Infinitorum*， 1655年）。使用负指数，并以符号∞表示无限。

牛顿 （Isaac Newton，1642—1727年）

英格兰人，物理学家、数学家、炼金术士。

将未知变量视为“流”（*fluent*，今日所称的“因变量”），亦即沿着曲线流动的量。以“打点”的形式来表示称为“标点”字母（*pricked* letter）的导数$\dot{x}$，$\dot{y}$，$\dot{z}$。

莱布尼兹 （Gottfried Leibniz，1646—1716年）

德国人，数学家、哲学家。

理解符号的限制及其在概念上的力量。为书写清晰起见，优先使用符号。发明特有的微分符号和积分符号。

欧拉 （Leonhard Euler，1707—1783年）

瑞士人，物理学家、数学家。

著有《皇家科学院报告编辑》（*Recueil des pieces qui ont remporte les pris de l'academie royale des sciences*，1777年），以i来表示$\sqrt{-1}$。

威廉·琼斯（William Jones，1746—1794年）

韦尔斯人，语文学家、古印度学家。

引进希腊字母π。

狄利克雷（Gustave-Peter Lejeune Dirichlet，1805—1859年）

德国人，数学家。

提出现代的函数概念。

哈密尔顿（William Rowan Hamilton，1805—1865年）

爱尔兰人，物理学家、数学家。

引进“四元数”，一种在四维中包含复数系的新数系。

第9章 欧几里得的秘密

许多年前，我拥有很难得的短暂时刻，获准翻阅现存最早的欧几里得《几何原本》复本“MS D’Orville 301”。这种少数人享有的尊荣特权，可以说不亚于获准进入女王的会客厅觐见女王。首先，我必须获得一位备受敬重的数学教授的推荐函。从封爵的教授那里获得推荐函或许不是必要的条件，但我的确得到了这样的推荐。然后，在预约那天，一名男士在牛津博德利图书馆特藏室外面的门廊上迎接我。这个瘦削的男子面容看起来像林肯一样，眉毛浓密，脸颊凹陷，他护送我进入一个房间，让我在那里进行宣誓程序。

Do fidem me nullum librum vel instrumentum…

> 我特此保证，不会拿走图书馆的东西，不会对属于图书馆或在图书馆管理下的任何书卷、文献或其他对象以任何方式标记、污损或破坏，不会带任何火源进入图书馆或故意引燃它们，也不会在图书馆里抽烟；而且我承诺遵守图书馆的所有规定。

在那里，我立誓尊重博德利图书馆的所有物品，并同意不做出清单里一长串的违规行为——不管是用笔、相机还是火。接着，我拿到一双白手套，并且按照要求用一支特殊的笔在欧几里得“MS D’Orville 301”的访客簿上签名。我看了一眼刚签完名的那一页，不由得怔了一下，因为我突然意识到我的签名很可能在那一页上留存千年，并且只在牛顿的签名下面十二行。

那个瘦削的男子突然将笔收回，以林肯般的严肃表情提醒我：“无论在任何情况下，请勿脱下手套碰触书页！”

我和这部伟大的文献单独留在房间里。与这样一部古代手稿独处的震惊、兴奋和荣耀之情，我实在无法形容。我就像是在中世纪修道院里的僧侣，像是在自家波希米亚图书馆里的伯爵，也像牛顿一样细细思索着为什么这份抄本里没有符号这个问题。单独在这个房间里，我觉得自己在精神上与上一个千禧年的所有昔日学者、抄写员和数学家有了一种关联，特别是与书记员史蒂芬（Stephan the Clerk），他在888年戮力为帕特拉斯的阿力他（Arethas of Patras）将这部作品抄写到羊皮纸上。我戴着白手套的手指小心地翻阅MS D’Orville 301的书页（参见图9-1）。

图9-1 博德利图书馆Big B的说明。资料来源：克雷数学研究所网站（http：//www.claymath.org/library/historical/euclid/images/euclid_1_48.jpg）。博德利图书馆，牛津大学，MS. D’Orville 301， fols. 31v-32r

当然，除了以一个字母表示点、两个字母表示线、三个字母表示角之外，书中没有其他数学符号。再者，当然，书中包含以连续的希腊字母系统来符号化的整数和有理数。但除了那些和我一样签下相同保证书的众多已作古的读者在书页边缘空白处的潦草涂写之外，我没发现任何代表加、乘或等于的符号。边缘空白处只填满了草草记下的印度数字，还有我幻想可能是牛顿所留下的几何涂写甚至代数方程式。

今天，我们也可以在网上浏览这份文献，不需推荐函，不必戴手套。这得感谢克雷数学研究所和博德利图书馆，让所有人都可以在线浏览完整的原版MS D’Orville 301。此外，每一本书都有内文图片的希腊文索引，附有英文翻译。

MS D’Orville 301展示了如何证明简单的等式，例如（$a+b$）$^2=a^2+b^2+2ab$；然而，你不会在欧几里得的著作中发现任何表示乘幂或加减的代数符号，因为他的著作是几何式的，而且完全用文字表述。即使是《几何原本》的首印版，里面也没有符号。数学史家托马斯·希斯英译本第二卷的命题7描述如下：

> 如果一条直线随机地切成两段，直线上所张拓出来的正方形，会等于两个线段各自张拓出来的正方形之和，再加上由这两个线段所形成的长方形之两倍。

尽管这是几何式叙述，但我们可以将它看成下面这个我们熟悉的方程式：

$$(a+b)^2=a^2+b^2+2ab$$

数学，它并不总是今日我们所见的样子。它的严密性不总是依赖有限数量的精心设计的陈述，并借由逻辑法则与基本的假设关联。我们今日的西方数学，承继了巴比伦人和埃及人千年来计算方法的具体应用成果，当时他们对数学证明的概念非常宽松。他们的主要目的是说服，而非严密性——严密性会在接下来的三百年里缓缓发展，而这些都发生在欧几里得和亚历山大学派基于基本的假设构造出证明的理念之前。

现存最早的几何史著作之一是普罗克洛斯的《欧几里得几何原本第一卷评注》。普罗克洛斯是哲学家、史学家，他总结了更早期由罗德斯的欧德谟所述的历史。5世纪之前的几位史学家，从普罗克洛斯到普鲁塔克，都告诉我们公元前6世纪的哲学家米利都的泰利斯将一种新的堪称奇迹的智慧引进希腊哲学：抽象几何学。

对于几何学之父，我们所知甚少。甚至欧几里得的生卒年也非常不确定。他可能是称为《几何原本》的这部巨著的编纂者和组织者，这部公元前300年的教科书概括了当时几乎所有已知的数学成果，但可以确定的是，书中的伟大证明不是这位数学家一人完成的。我们说“欧几里得证明了……”，只是指这个那个定理出现在《几何原本》的十三卷当中。更有可能的情况是，欧几里得这位老兄是从其他与柏拉图学院有关联的人那里学习到书中的许多定理的，诸如欧多克索斯（Eudoxus）和泰提特斯（Theaetetus）等人。若不使用某种公理化的逻辑方法来证明数学定理，那么，学习这些定理便不能说是证明了它们。因此，尽管某些几何定理可以凭借可信的论证直观地被认定为真，但是毫无疑问，这些证明比起欧几里得所建立的辉煌体系，是不太有说服力的。欧几里得是从公理或不证自明的真理出发，并遵守无可争辩的逻辑法则来建立数学证明。《几何原本》赋予了数学基本特性，它是**证明**的第一个模型。

在欧几里得的《几何原本》问世之前数百年，人们便已知晓勾股定理。埃及人知道它，中国人知道它，印度人知道它，毕达哥拉斯学派成员当然也知道它。但是，它的证明无疑出现在《几何原本》第一卷之末，紧跟在前面四十六个命题之后。

我们也知道欧几里得这位老兄，在亚历山大大帝于公元前331年建立亚历山大城之后不久在那里生活。亚历山大大帝将这座城市交由他的将军托勒密一世统治，并由狄诺克拉底规划整座城市的建设蓝图。这座城市被建造成坐标式的方格状，街道相互垂直。到了欧几里得时代，整座城市正处于蓬勃发展时期。城里的两座大图书馆，已经收藏了从进入港口的船只那里征收的书籍。剧院中上演着一出出希腊悲剧，教授哲学的学校急剧增长。

欧几里得的时代之后五百年间，亚历山大城始终是学习和钻研数学、科学、医学的中心。2世纪，时值丢番图生活的时代，这座城市仍旧充满了惊奇。城中宽阔的林荫大道和石板小路，夜晚被火炬照亮，比此后两千年内欧洲任何一座城市的夜晚更明亮。亚历山大城中的石

灰岩柱廊依然可见，从城墙的一侧延伸到另一侧。城里还有几座公园，以及克丽奥佩特拉的纪念碑。香水和羊皮纸，还有玻璃工艺和雪花石膏雕刻，仍是这座城市蓬勃发展的制造业当中重要的一环。城里还有神殿和犹太会堂。城中还有许多摊贩、表演者、放贷者和娼妓。这是“一座极具感官享受、气象万千并充满高度人文气息的城市，古代世界的巴黎”。难怪这么多数学家来到这里工作。

丢番图也许诞生于此，虽然我们无法确定，但我们可以肯定的是，他的主要研究，尽管在中世纪的黑暗时代完全被人遗忘[1]，但当那些成果于16世纪重现时，深深影响了代数学的发展。

在那个年代，“阅读”（reading）仍指的是大声读出来。默读在当时并不存在，即使在公开场合亦然。阅读是一件专心致志的事，需要随时滚动并平衡卷轴。尽管到了丢番图的时代，字与字之间已经分开，但是早期的手稿并未将词、句、段分开，也没有标点符号。阅读本身是一件困难的事，并且被视为了不起的能力。读与写多半是在晨间明亮之时进行，特别是还不太热的时候。丢番图可能是用海边的灯芯草秆在羊皮纸上书写的。

如果我们相信抄写员和翻译者忠实抄写和翻译了丢番图的原稿，那么丢番图的确使用了符号来表示乘幂和未知数。以未知数来看，他使用了类似希腊字母sigma置于词尾时的写法[2]，尽管它比一般文本中所用的sigma更大且更加倾斜。它可能是“数”这个词的古希腊文前两个字母的草书体缩写。也许它并非我们现在看到的样子，但著名的古代数学史家托马斯·希斯认为，它可能是在日益严重的马虎抄写过程中慢慢演变的结果。丢番图也创造出减号，它就像一支朝上的箭（有时是朝下）。但这也是一种缩写形式——很可能是“减”这个词的前两个字母，同时，其中一个套着另一个。或者，那可能只是后世的某位抄写员自己想出来的，利用这些符号当作缩写。

如果所有的数学完全以文字表述，不使用符号或没有大量精心设计的符号，那么，这门学问看起来会是什么样子？下面是阿尔-花拉子密《代数》中的一段，甚至连数字也以文字来呈现：

> 当一个正方形的量加上二十一个迪拉姆（dirhem），与那个正方形之根数的十倍相等，则此正方形之量为何？

我们可以将这个问题简写为$x^2+21=10x$。

该书中所给的解并不包含符号，如下所述：

> 将根数折半，得其应有部分五。将其自乘，所得结果为二十五。减去平方数所加之数二十一，余为四。开平方，得到二。从根数的应有部分，亦即五，减去此数，所得为三，这即为你所求的正方形之根，其平方数则为九……

需要继续说下去？

这个问题来自一个稍微更具应用性的问题：

> 我将十分成两个部分。接着，我将其中一个部分乘上另一部分，所得结果为二十一。那么，你现在可以知道，十所分成的两个部分，其中的一部分为何。

阿尔-花拉子密解答问题的语言，给我们提供了一个似乎专属于该问题的逻辑过程。它也许是一种常见的方法，一种隐藏在文字表述背后的算法，但可能需要花些功夫才能表示这个过程。另一方面，符号化的代数过程，提炼出了很多这类问题的解答。以现代的符号化术语来表示，问题的解答会像这样：十被分成两个部分，其中一部分可能比另一部分大。所以，这两个部分可分别表示成x和$10-x$。这两个部分的乘积必定等于21。因此，$x(10-x)=21$。由此可得二次方程式$x^2-10x+21=0$，而其解为$x=3$或$x=7$。[3]

然而，请记住，这个问题不像是通过文字表述的方式解出来的。它可能一开始是在某种沙板上做研究，接着再形成文字表述。再者，如何将10分成两部分，使得其乘积为21的这个问题，有一个利用心算的简单解法：21的因子中，只有3和7这两个因子满足其和为10。这也是标准的巴比伦式几何问题：列出两个部分x和y，接着，考察这个问题，x和y分别为一长方形的两边，且满足其和为10，面积为21。这个几何问题可以用下面涉及两个方程式的代数问题表示：

$$\begin{cases} x+y=10 \\ xy=21 \end{cases}$$

将第一式当中的y用x表示，代入第二式后可得$x^2-10x+21=0$。

阿尔-花拉子密的证明是几何式的，并不是我们称之为代数的现代意义的代数解法。这不让人意外，在阿尔-花拉子密的时代，阿拉伯数学文献里并没有代数证明。但是，在没有任何说明的情况下，阿尔-花拉子密的确提出了一种以文字表达的算法，可以从问题的条件逐步导出答案。利用很难理解的文字代数，他告诉我们如何进行下去。不过，当时没有符号，一个也没有，甚至没有数字。数学史家贺鲁伯告诉我们：“与解基本方程式相关的代数式证明，在整个阿拉伯传统中缺席……我们应该预期，一种代数符号体系的存在，与由于后见之明而看起来似乎可能创造出的这种推理，这两者之间没有直接的关系。”

我们不应愚蠢地认为，以文字表达的符号形式，只是一种方便的速记法。它当然是速记法，但不仅如此，它还有助于人类思维超越自然语言所书写成的文字所带来的含混不清与误解。更进一步，这种符号体系允许思维将特殊的陈述提升至其一般化的形式。到了笛卡儿的时代，方程式几乎完全以现代的符号形式写成，符号终于达到了——如托比亚斯·丹齐克所言：“从文字的奴役中，将代数解放出来。”

---

[1] 黑暗时代（Dark Age）之说目前已经被史家摒弃了。

[2] 古希腊字母sigma一般大写为Σ，小写为σ，但置于词尾的sigma 小写字母则写成ς。——译者注

[3] 根据二次方程式公式，其解计算如下：

$$\begin{aligned} x &= \frac{10}{2} \pm \sqrt{\left(\frac{10}{2}\right)^2 - 21} \\ &= 5 \pm \sqrt{25-21} \\ &= 5 \pm \sqrt{4} \\ &= 5 \pm 2 \\ &= 3 \text{ 或 } 7 \end{aligned}$$

第10章 讽刺短诗式的谜题

当你可以理解它们之后，方程式其实相当友善。

——伊恩·史都华（Ian Stewart）

根据叙利亚的哲学家杨布里科斯的说法，可称为代数的古代最早一批著作，可追溯至早期毕达哥拉斯学派，至少可能追溯至该学派的帕罗斯的塞马里达斯，其中提出求解某类包含n个未知数的n个联立方程式的法则。以三个未知数为例，法则可简化为：给定三个量之和，以及三个量当中某一指定量与其余各量配对之和，则指定的量会等于这些配对之和与三量之和的差。

按照我们现代的符号语言，这个问题可以更简化如下：

如有下列联立方程式

$$\left\{\begin{array}{c} x+y+z=a \\ x+y=b \\ x+z=c \end{array}\right\}$$

则$x=b+c-a$。

举例来说，如有下列联立方程式

$$\begin{cases} x+y+z=3 \\ x+y=2 \\ x+z=4 \end{cases}$$

则$x=2+4-3=3$。

这是一个简单的代入消去过程，但基本上，它是自19世纪后便为大家熟知的克拉玛公式（Cramer’s rule）。在塞马里达斯的时代，它被称为“塞马里达斯之花”（the flower of Thymaridas）。如果我们继续求其他未知数，我们会得到$y=-1$及$z=1$。尽管解$y=-1$行得通，但它却被认为是荒谬的，因为-1是一个负的量。分数和有理数都可以被接受；然而，16世纪之前，负数——它被视为负债当然没问题——在欧洲不被视为一个真实的数。[1]

欧几里得《几何原本》（MS D’Orville 301）9世纪版的第二卷、第五卷和第七卷，都无意间利用几何的语言处理代数，也就是说量被刻画成给定的线段长，成为譬如二次方程式$x^2+ax=b^2$（用我们现

代的记法）的解，其中a和b是正数。对身处21世纪的我们而言，它是两个代入x后，可以使得方程式成立的数；以其中任一个数代入计算，可使得等号的左边与右边相等。但对生活在15世纪之前的人来说，方程式的解是在没有符号的情况下求得的，同时，他们还依循了什么**东西**可被视为数的一种具体观点。

方程式$x^2+3x=4$有两个解（$x=1$和$x=-4$），然而，只有一个是正的，因此，只有一个解被视为数。这类方程式可以用几何的方式求解，事实上，它可能是因为实用的几何问题而得到的灵感，比如说，试求一个长比宽多三个“斯塔德”（*stadion*，1斯塔德约185米），且面积等于四个平方“斯塔德”的长方形的宽。在这个例子中，对这样的几何问题，满足方程式$x^2+3x=4$的负数解-4，似乎无法应用于边长应为正的长方形。15世纪的数学家无法知道，几何本身何以包含满足该方程式的负数。方程式本身给了几何没有找到的某些东西，即便负数解在几何学里亦表示了相当真实的东西。

现代数学来自三个基本的根源：代数、几何和分析，而逻辑被视为三者共同的基础。这些根源盘根错节地纠缠在一起，也因此，我们要区分出这些领域的根源是很困难的事。我们现在有代数几何学，一个相当新颖的数学分支，结合了抽象代数与几何学的技术；也有几何分析学，一个使用几何方法来研究偏微分方程的学科；还有解析数论，数论的一个分支，使用分析的方法来解整数的问题。不过，基本上，在数学非常古老的树根处，我们发现了代数、几何和分析。

符号化的现代数学，最基本的形态可以追溯到丢番图的《算术》一书。我们应提醒读者，该书的原始文本现已不存，所以在复本上发现的任何记法，都可能是抄写员或翻译者添加的。

丢番图撰述《算术》的年代，晚于亚历山大的海普西克利斯（因为丢番图引述了海普西克利斯的话），并且早于希帕蒂亚的父亲亚历山大的塞翁（因为塞翁引述了丢番图的话）。这让我们能够推定他生活的年代，大约介于公元120年至400年间。另外，有一封11世纪拜占庭僧侣的信件宣称，老底嘉（Laodicea，位于今土耳其西南部）主教亚纳多留斯在250年左右，题献了一部专著给丢番图。因为这个题献的记述，后人推测丢番图活跃的年代应该不会晚于250年太久。

然而，20世纪80年代，著名的数学史家韦尔伯·柯诺怀疑，有一本一直被认为是亚历山大的希罗（Heron）著作的书，应该是出自丢番

图。柯诺研究了那本据称是希罗的著作的风格，发现该书与丢番图的著作风格非常相似。他推测老底嘉主教信中所提的丢番图另有其人。由于希罗死于公元70年，所以《算术》的成书年代被认为较可能是1世纪，而非3世纪。

一首墓志诗或许让我们知道丢番图活了多少年：

> “丢番图沉睡于此”，好奇者注视。
>
> 通过代数的技术，墓碑诉说着年纪：
>
> “上帝赐予他的童年时光，占了一生的六分之一；
>
> 再过了十二分之一的年少岁月，胡子爬上双颊；
>
> 七分之一的时间过去，他结了婚；
>
> 五年后，喜获麟儿。
>
> 可恨啊，大师与圣者所挚爱的这个孩子
>
> 仅活了父亲年纪之半，冷酷的命运将他带走。
>
> 在利用数字科学慰藉他的命运四年后，他告别了这一生。”

这个代数谜题摘自7世纪希腊文集《帕拉蒂尼文集》，作者是梅特罗多勒斯。在当时若要求得答案，需要精通《算术》一书中的知识；但利用我们现代的符号代数，可以很快求得其解。

依照这首诗的描述，我们发现丢番图的童年占其一生的 $\frac{1}{6}$。又经过人生的$\frac{1}{12}$后，他长出了胡子。再过$\frac{1}{7}$，他结了婚。五年后，他的儿子出生，儿子仅活了父亲年纪的$\frac{1}{2}$时间。儿子去世后四年，丢番图也走完了人生的旅途。所以，如果我们假设丢番图活了x岁，而他的儿子活了y岁，那么我们知道

$$x = \left(\frac{1}{6} + \frac{1}{12} + \frac{1}{7} \right) x + 5 + y + 4$$

又已知

$$y = \frac{x}{2}$$

上述可视为包含两个未知数的联立方程式，但它们可以迅速简化成仅带有一个未知数的方程式。利用代入消去法，将第二式的y代入第一式，我们可求得丢番图去世时是八十四岁。多么简单呀！

《帕拉蒂尼文集》收录了四十六则讽刺短诗式谜题，其中许多可以导出简单的联立方程式代数问题，而这些问题都源自一则几个人分苹果的传统问题。这类代数谜题，可回溯至公元前5世纪之前，它们都未使用任何符号。举例来说，已知有六个人分若干个苹果，其中，第一个人得到全部的三分之一，第二个人得到八分之一，第三个人得到四分之一，第四个人得到五分之一，第五个人得到十个苹果，第六个人只得到一个。请问苹果共有几个？

我们可以把它看成一个求x的问题：

$$\frac{1}{3}x + \frac{1}{8}x + \frac{1}{4}x + \frac{1}{5}x + 10 + 1 = x$$

你知道x是多少吗？借由我们的符号代数工具，我们可以解这个方程式。合并同类项，同时从等号两边减去x，很快求得答案：x＝120个苹果。

就如同一种规律，希腊著作历经多个阶段，先是被翻译成叙利亚文，再译成阿拉伯文，接着被译成拉丁文，每一个阶段多少增添了不正确之处。其间，译版也辗转历经波斯文、叙利亚文、阿拉伯文、阿拉姆文和其他语种。阿拉伯人对于科学、数学、力学和哲学最感兴趣——阿波罗尼斯、斐洛、阿基米德、希罗、柏拉图、亚里士多德、泰奥弗拉斯托斯。到了9世纪中叶，巴格达、拜占庭和地中海东岸其他城

市，由于学术风气日盛，翻译的需求增加。在巴格达，有一位侯奈因·伊本·伊斯哈格，他十七岁就精通多国语言，创办了一所翻译学校。侯奈因怀疑许多希腊手稿散落在伊斯兰世界各地，亲自带队前往美索不达米亚、叙利亚和亚历山大城找寻这些手稿。他轻视早期的翻译者，认为那些人要么完全无法胜任，要么浪费时间在损坏或难以辨认的手稿上。

侯奈因创办的学校很特别，因为至少从现代语文学的标准来看，他的翻译技巧与众不同又正确。他的学校教导学生小心翼翼地比较他们所能发现的不同时间、不同出处的相异版本手稿。“感谢侯奈因和他所带领的学者，许多希腊著作以高质量的阿拉伯文译本留存下来。”

4世纪之前，书籍中的文字是以安色尔字体（uncial，大写字母）写成。虽然在接下来几个世纪，出现一些实验性的小写字体，但在侯奈因的学校创建之前，都没有什么太大的改变。安色尔字体最主要的缺点是写起来太慢，字又太大，每一页的字数受到很大限制。为了降低昂贵的书写材料成本，小写字体（minuscule，用于书信和官方文件的小写字母）取代了安色尔字体。新的字体使得抄写变得更加容易又便宜，书籍可以更快速地被抄写，但文本内容中模棱两可字体所造成的麻烦，还是需要译者的诠释来解决。

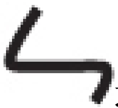
阿拉伯人于641年征服了埃及之后，羊皮纸的需求剧增，即便当时的人对文献并没有太多兴趣。莎草种植园被耗尽，书写用的材料不再价廉易得。但到了850年，伴随着学术的复苏——或者很可能就是因为复苏之助——手稿在外观和制作方式上皆出现了变化。

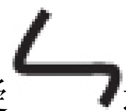
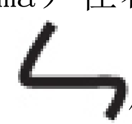
接着，751年，在塔拉斯之战（Battle of Talas）中，阿拉伯人成功阻止了中国人往西扩张到哈萨克斯坦，而在战争过程中，两名中国战俘被带到乌兹别克的撒马尔罕。哈萨克斯坦的阿拉伯人从这两名中国士兵身上，学会了造纸术。纸张使得书写变得便宜。因此，9世纪时，以前的安色尔字体文本被译写为新的小写字体，从而保存了希腊最精华的文献。古希腊著作所有后来的抄写本，前身都是一个或多个以安色尔字体写在莎草纸上的版本；几乎所有这些文本都可溯源至它们9世纪的范本。

不幸的是，译写过程中出现一些错误，字母产生混淆或误读。希腊文本中许多错误通常都源自相同来源的9世纪手稿抄写本。从安色尔

字体转抄成小写字体之后，原始的版本被丢弃，而小写字体抄写本变成所有日后抄写本的参考版本。所以，许多古文本后来只留下唯一的抄写本。例如丢番图的《算术》，据其序文可知原始版本包含十三卷，仅有其中六卷和第七卷的一部分留存下来。

快速一瞥《算术》这本书，即能发现它所蕴涵的代数特色。这也是为什么过去有些史学家认为代数学起源于丢番图。更仔细地扫视，则会发现它的杰出成就和记法的生硬。这本书教导读者如何解特殊的一次方程式与二次方程式，但它的记法看起来似乎是由那些用来计算求解的未知数和乘幂的缩写所构成的。

在该书中，对某位名为戴奥尼修斯（Dionysius，无关酒神的另一位）的人士，提出关于平方、立方和数字的其他一般性质的难题并求解时，作者根据字母来为平方、立方定义名称，并将未知量表示为ἀριθμός，意指“该数”。没几页之后，丢番图就用了这个符号，好像是因为用整个词ἀριθμός太麻烦。

两百年间，学者始终质疑这个符号的起源。有些人认为它是希腊字母sigma里只写在词尾的形式ς。这种想法是基于丢番图知道在ς与一个数之间不会产生混淆；希腊字母表的字母具有数值上的等价物，但这个最后的字母ς（另类的sigma）在希腊的数字系统中从未被视为一个数。持这种看法的学者认为，仅表示大写且倾斜的ς。然而，另一种观点偏向于认为表示

ἀριθμός前两个字母ἀρ（第一个音节）的速记简写，而根据我们对符号的定义，它根本不可能是一个代数符号。

所有可贵的论点都有值得参考之处。进入20世纪，著名数学史家托马斯·希斯提出了具有说服力的理由，他认为ς既不是词尾的sigma也不是某种象形文字，而是ἀριθμός这个词前两个字母的变形。他推论，它建立了一种“丢番图所使用的不同缩写之间的一致性。这表示他一直使用一个不变的原则来设定那些缩写，而我们会因为预期他有一贯的规则而易于理解”。对应到希腊文前几个字母的μ，δ和κ，它们分别用来表示单子（monad，我们的未知数*x*）、平方和立方。希斯认为这些字母也可能与它们所对应的数值等价物产生混淆——也就是40，4和20。为了避免这样的混淆，对于μονάδων，δύναμις和κύβος这三个希腊词，丢番图必定得加入每个词的第二个字母。但这时所形成的μο，δύ和κύ[2]，会分别与数值等价物4070，4400，20400产生混淆。为了避免**那种**混淆，每一个字的第二个字母写成上标形式，缩写分别变成了μ^{o}，$\delta^{\acute{\upsilon}}$和$\kappa^{\acute{\upsilon}}$。应用在ἀριθμός这个词时，缩写成为ἀ$^{\rho}$。偶然地，其他符号也出现了，诸如$\mu^{\bar{o}}$，它代表的是一个不确定的数。

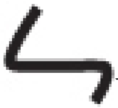
从ἀ$^{\rho}$演变成ς的过程，希斯又做何解释呢？现在，我们先忽略下述事实，在丢番图的时代，分数的分母一般写在分子的右上角，这意味着ἀ$^{\rho}$可能与表示$\frac{1}{100}$的符号产生混淆。

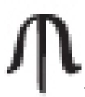
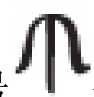
抄写员以草写体抄写的过程中未必总是小心翼翼。抄写员长时间的快速工作——有时靠着窗边朦胧的光，有时在幽暗微弱的烛光下

——很可能使得草写的两个希腊字母α ρ，被抄写成类似ϛ或S的形状，因为这两个S形字母都用在后来的翻译中。希斯提醒大家留意19世纪著名语文学家维克托·埃米尔·卡德豪森（Viktor Emil Gardthausen）的说法，后者论述说古代手稿中的草写文字历经了不同阶段。两个希腊字母ἀρ演变成为υρ，也许是因为丢番图用它来速记ἀριθμός这个词。接着，根据希斯的说法，经过一代又一代的抄写与再抄写之后，抄写员不再把它看作两个字母，而是抄写成我们所见的某种含混不清的小写字母形式。

抄写员的工作是抄写，而非改写编辑，可以肯定的是，他们的工作不是增加或改变著作的内容。抄写员不是僧侣就是受雇的专业人员，他们对于自己所抄写的内容往往全然不懂，特别是当他们抄写的是关于科学或数学的书籍时更是如此。整个工作过程无人理会，他们得独处好几个月，有时一次就好几年。他们所抄写的书籍的作者，往往已去世多年，甚至几百年，所以找不到足够权威的人来检查错误。那些抄写员自行润饰、添加、删除，并犯下错误。由于像《算术》这样较著名的著作通常不是原稿，而是复本，所以错误更加严重，以至于中世纪史专家尝试区别原著与抄本的差异时，常常被激怒。

希斯的论述在我看来表面上讲得通，许多学者质疑过这个看法。20世纪早期的苏格兰数学生物学家达西·温特沃斯·汤普森对ϛ的成因提出自己的论点。这个符号通常写成带有附加物的弯曲形式，如ϛ′或ϛοῡ，或者（复数形式）ϛϛοι′，这表明该符号被当作一个单词的一部分。19世纪数学史家詹姆斯·高在希斯的想法出现于20世纪早期的数学史圈之后不久，写了一篇通俗易懂的文章。他认为这既不是前两个字母的缩写，也不是ἀριθμός这个词词尾的sigma。由于ϛ′仅出现在草写的希腊文里，而草写的希腊文在

8世纪之前尚未出现，因此他排除了词尾的sigma这个选项。由于怀疑可能来自某种讹误的速记，他欣然接受它很可能来自印度文或巴比伦文字或僧侣体（埃及的草书）的想法。他的友人塞缪尔·伯奇是埃及古文物学家，伯奇告诉他，在形式上与纸草书中的一个僧侣体符号几乎一样，这个符号代表一种未知的力量，也代表一“堆”（古埃及文*hau*）。阿美斯（Ahmes）利用纸草书中的一个僧侣体符号来表示未知数，他是著名的莱因德纸草书的抄写员。这本实用问题手册约成书于公元前1550年，“是了解事物的精确计算及现存所有事物知识的指南”，现存于大英博物馆。所有的僧侣体符号在形式上有些微差异，并且衍生自不同的象形图案；然而，代表“总和”的符号看起来也很像那个纸草书上的符号。詹姆斯·高发表了他的论述之后，希斯提出反驳。因此，整个问题仍悬而未决。

丢番图的确没有使用任何代表“加”的符号。然而，这里存在另一个难解之谜：1621年，法国数学家、语言学家、学者克劳德·加斯帕尔·巴歇将丢番图的《算术》翻译为拉丁文。根据这个译本，丢番图清楚告诉我们（卷一，定义十一）：“‘缺乏’（*wanting*）乘以‘缺乏’会得到‘给予’（*forthcoming*），而‘缺乏’是用字母ψ倒过来之后截去顶端形成的来表示。”通常符号会用来表示减，它是一个真正的符号，它是抽象的，与单词“减”（minus）没有显而易见的直接关联。

DEFINITIO IX Minus per minus multiplicatum, producit Plus. At minus per plus multiplicatum, producit minus. Et defectus nota est litera ψ decurtata, & deorsum, sic ⋔.

翻译：定义十一：以少乘少则得多，以多乘少则得少。而减则可以利用截短字母ψ再上下颠倒来表示。

这里，我们找到代表减的符号的第一个证据。丢番图告诉我们，他用来表示减的符号⋔来自希腊字母ψ，将这个字母的尾巴去掉，再上下颠倒。然而，书中的记法并不完全一致，有时用这个符号，有时在希腊文原本里用λείψει（缺乏）这个词，甚至同一页出现不同的符号。

亚历山大的希罗的著作《测量术》（*Metrica*）中同样出现了这个符号，该书写于1世纪，这意味着在丢番图出生之前这个符号便已在使用了。它很可能是λείψει这个词的一种缩写，或许是取第一个字母与最后一个字母的组合，或者可能是某个僧侣体文字。⋔似乎是丢番图所用的唯一未与书写的文字直接相关的真正符号。《算术》中的所有其他标记看起来都是缩写形式。因为所有现存的复本都可追溯至13世纪，很难知道是谁使这些符号经历了这样漫长的岁月。

要表示两项之和，丢番图（或《算术》的抄写员）仅是将它们并列。举例来说，把幺元$\overset{\circ}{\mu}\,\overline{\alpha}$与未知数$\varsigma^{\overline{o\upsilon}}$接在一起形成$\varsigma^{\overline{o\upsilon}}\ \overset{\circ}{\mu}\,\overline{\alpha}$，用来表示$x+1$这个多项式，或者再简化成$\varsigma\overset{\circ}{\mu}\,\overline{\alpha}$。然而，他能利用交换加、减的顺序，以及合并同类项，来简化方程式，就像我们的做法。所有过程都通过文字的方式，不需要任何程序法则。他假设我们必定从其他地方知道了一些运算法则，也许是从其他书籍或其他教师那里习得的。所以，当他神乎其技地求得了问题的解，他便停了下来。附带一提，我们写带分数时还是用并列的方式，亦即$2\frac{1}{2}$指的是$2+\frac{1}{2}$。

《算术》的抄本

希腊文原版的十三卷里，只有六卷以抄本的形式留存下来，十三卷中的四卷（卷四至卷七）阿拉伯文版直到近期的1968年才发现。现在几乎所有对《算术》这本书的批注，都是来自巴歇的拉丁文译本，这个名为“Parsinius 2379”的翻译手稿，是在1545年之后，由海德隆提乌斯抄写的。这份手稿现藏于巴黎的法国国家图书馆，是包含希腊文原文的最早版本。追溯《算术》的起源很困难。现存最早的手稿年代晚于保存不善的13世纪版本“Matritensis 48”，现藏于马德里国家图书馆。

欧洲的大型图书馆起源于小型研究室。最早的大学当中有些设立在意大利的小城市——波隆那、佛罗伦萨、那不勒斯、帕多瓦、帕维亚、佩鲁贾、比萨、罗马、锡耶纳，时间早于13世纪中叶。早在梵蒂冈图书馆出现之前，这些大学就已成为意大利的学术中心。对许多这样的城市来说，一所大学仍只是学生接受不同教师的学术教导而结合在一起的组织，它并非实体机构。全欧洲富有家庭的学生来到这些意大利小镇，以当时通用的语言拉丁文来学习，他们直接付费给授课的教师。自由，从职业教育中解放出来，他们是接受博雅教育的学生。

1463年，德国数学家兼天文学家约翰尼斯·穆勒，拉丁名为雷吉蒙塔努斯，应邀到帕多瓦大学演讲，当时这所学校已创建两百多年。在那些相关讲座中，他介绍了所有的数学科学。“没有人，”他说道，“曾将丢番图的十三卷杰作从希腊文翻译成拉丁文，使得算术美丽的花朵被隐藏起来，那*ars rei et census*，今日他们以阿拉伯名称之为代数。”[3]这很可能是欧洲的作者第一次提到丢番图的著作。接着，在一封写给意大利数学家乔瓦尼·比安契尼的信中，他写到他在威尼斯发现“希腊数学家丢番图的著作未曾被翻译成拉丁文”。似乎没有人确知雷吉蒙塔努斯是如何发现《算术》的这个抄本的。近1620年的某个时间，巴歇声称裴隆枢机主教拥有一份包含了丢番图完整十三卷著作的手稿。根据裴隆的说法，这份手稿借给了一个朋友，而在取回手稿之前，那个朋友就去世了。

在奥斯曼帝国于1453年围攻君士坦丁堡之前两个世纪，一场大火烧毁了城里一座大图书馆的十万多册藏书。但几年后，这座图书馆竭尽全力，将希腊文和阿拉姆文著作翻译成阿拉伯文，并且雇用数百名抄写员，将碎裂的纸草书古本重新抄写到羊皮纸上。由于某种原因，

那些曾属于君士坦丁堡大图书馆的抄本成为战利品被带到西方，最终落入私人收藏家、全欧洲不断发展的大学图书馆，以及梵蒂冈之手。

接着在1448年，教皇尼古拉五世在教皇宫殿设计了一座公共图书馆。它一开始是个有大窗和壁画的房间。教皇认为极重要或有精美装饰的书籍，都被锁在长椅上。这座图书馆本身就是精美的艺术品。到了1455年尼古拉五世去世时，这座图书馆有一千多本藏书。1475年，教皇思道四世任命了第一位梵蒂冈图书馆管理员巴尔托洛梅奥·普拉提纳，他手写了一份三千五百项的目录，收录的藏书是欧洲最多的。

这座图书馆收藏的多是神学书。然而，到了普拉提纳六年任期结束时，以希腊文和拉丁文写成的世俗著作藏书，已使这里发展成为西方世界收藏经典著作最重要的学术中心。这里藏有数千册关于艺术、音乐、哲学、神学、罗马教会史、科学和数学的装饰华丽的手稿，这些手稿是从远至东方的中国等王国和各帝国那里购买或掠夺而来。我们现在知道，当时梵蒂冈图书馆有至少两部丢番图著作的抄本。

16世纪的德国学者克胥兰德告诉我们，1571年10月，他偶然发现了一部《算术》的抄本。当时他在威登堡和几位数学家聊天，那几位数学家已经有《算术》手稿中的几页，而那份手稿是一位叫安德鲁斯·都狄修斯的人所有。克胥兰德离开威登堡前往莱比锡之前，抄写了其中一个问题及其解答，并给莱比锡的一位教授西蒙·卢森希斯看，后者回信说希望看看那份手稿。

年代次之的手稿是15世纪的Matritensis 48抄本（Vat. gr. 191），收藏于梵蒂冈教廷图书馆。三个世纪后，法国数学家兼数学史家保罗·坦纳利编纂整理了一份清单，列出了13世纪至16世纪的二十三部《算术》抄本。

生活在5世纪的希帕蒂亚曾有一部后来散佚的抄本。有参考文献指出，8世纪或9世纪曾有《算术》的抄本。从丢番图撰写《算术》到抄本Matritensis 48，历经近千年时间。传抄再传抄，从希腊文到阿拉伯文，到阿拉姆文，又回到希腊文，必定不仅出现错误，也添加了一些额外的内容。我们现在归功于丢番图原作的简字符号，会不会是以这种方式，出现在某一部抄本中的？

从一部抄本到另一部抄本里的记法，非常难以追踪。研究三种译本——希斯的英文版、克胥兰德的拉丁文版、巴歇的拉丁文／希腊文

对照版，我们发现译本中的差异毫无来由。希斯将书中的符号转写成某种形式，而在克胥兰德和巴歇这两个版本中都看不到那种符号形式，但几乎所有普遍流传的文献都采用希斯转写的版本。在马德里手稿（Matritensis 48）中，未知数表示成，非常像将拉丁字母h垂直与水平翻转。在15世纪的威尼斯手稿（Marcianus 308）中，看起来像S的同一个标记在博德利手稿里是。希斯认为，所有那些符号都只是一种缩写的讹误。这样的假设更符合从μονάδων，δύναμις和κύβος演变为分别以符号μ，δ和κ来表示平方（乘幂）、立方的过程的看法。[4]

甚至巴歇的译本中也出现好几种未知数（我们现代的*x*）的记法。在他的定义二中，出现了某个看起来像希腊字母ς的符号。有时它有重音符号ς′，有时带有上标ς″，有时上标上还有上标$\varsigma^{\text{ο}\prime}$。这些都是丢番图用来称ὁ ἀριθμός，也就是“该数字”这个词的速记写法。有时我们发现那个符号写成$\varsigma^{\text{οι}}$，有时则写成$\varsigma^{\text{όν}}$。这些差异反映了“数字”这个间接宾语使用时的语法或语义形式，因为它们反映出ἀριθμός（ἀριθμοί或ἀριθμόν）词尾的各种可能性，视其在句子中的语法结

构来决定。一个双sigma是指复数，在同一页里，我们会看到$\varsigma\varsigma^{\text{ούϛ}}$或$\varsigma\varsigma^{\iota o \nu}$或$\varsigma\varsigma^{\iota o \varsigma}$，同样视语法而定。

字母ς也出现在柏拉图学派哲学家士麦那的塞翁（Theon of Smyrna）的著作里。他生活在2世纪早期，所以他很可能是最早想到把缩写用在数学中的人。

依巴歇的记法，Parsinius 2379中的多项式译为$9Q+14-9N$（其中Q代表x^2，N代表x），而我们现代的记法则写为$x^2-9x+14$。请注意$\varsigma\varsigma\tilde{}$。丢番图用这个记号来表示复数，因为他将九个负项合并成一项。注意丢番图所用的系数（即$\vartheta=9$和$\iota\delta=14$）写在类别的后面（即$\varsigma\varsigma\tilde{}\vartheta$）。根据我们现代的记法，$\varsigma\varsigma\tilde{}\vartheta$指的是$9x$，其中$x$以双sigma表示，代表复数，也就是9倍的$x$。换句话说，$1x$可以写成$\varsigma\alpha$或只用$\varsigma$来表示，而$2x$可写成$\varsigma\varsigma\tilde{}\beta$（参见表10-1）。

表10-1 丢番图的记法表

μ°	幺元，例如 $\mu^{\circ\varepsilon}$ 表示 5 个幺元
⋔	减
ι^{σ}	等于，可能来自表示“相等”的词 ισος 的前两个字母
ϛ	未知数，即 x
δ^{υ}	平方，即 x^2
κ^{υ}	立方，即 x^3
$\delta^{\upsilon}\delta$	四次方，即 x^4
$\delta\kappa^{\upsilon}$	五次方，即 x^5
$\kappa\kappa^{\upsilon}$	六次方，即 x^6

备注：在古典批注中，这些符号都是大写。参见《亚历山大的丢番图全集》。亦参见希斯《亚历山大的丢番图：希腊数学史研究》，448 页

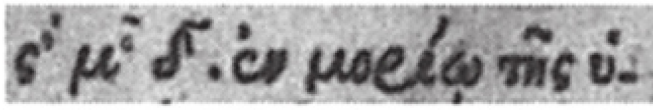
所有这些或许显示丢番图的确曾表示出未知数，但并非用一个显然不是单词的符号（但概念上仍与那个单词有关联），而仅是一种缩写形式。然而，19世纪数学史家保罗·坦纳利宣称，拜占庭时代之前的古代手稿并未使用这些不同的语法案例，可能是后来的抄写者自己擅自纳入了这些词尾缩写形式。如果丢番图对未知数 ἀριθμός 的不同语法词尾采用了不同的记法，那么为什么其他符号没有用同样的做法？希斯对于丢番图真的使用词尾的sigma作为代表未知数的缩写这样的论述表示怀疑，使得关于丢番图记法的猜想更加令人困惑。他推论词尾的sigma是后来才加入希腊字母表的。巴歇对定义九的翻译加深了这样的怀疑，因此认为词尾sigma与未知数的缩写无关。

翻译成我们现代的记法，丢番图的多项式 $\kappa^{\gamma}\overline{\gamma}\pitchfork\delta^{\gamma}\overline{\beta}\varsigma\overline{\alpha}\mu^{\bar{o}}\ \overline{\alpha}$ 会变成$3x^3-2x^2+x+1$：

$$\underbrace{\kappa^{\gamma}\overline{\gamma}}_{3x^{3}}\quad\underbrace{\pitchfork}_{-}\quad\underbrace{\delta^{\gamma}\overline{\beta}}_{2x^{2}}\quad+\quad\underbrace{\varsigma\overline{\alpha}}_{x}\quad+\quad\underbrace{\mu^{\bar{o}}\overline{\alpha}}_{1}$$

丢番图用X这个标记来写倒数。要写$\frac{1}{x}$，他会写成ς^{χ}。但除法可以用 ἐν μορίῳ 来表示，它指的是“分享”。所以用我们现代的记法，$\delta^{\upsilon}\ \tau\mu^{o}\ \alpha\psi\kappa\varepsilon$ ἐν μορίῳ $\delta^{\upsilon}\ \delta\alpha\mu^{o}\ \pi\delta\ \delta^{\upsilon}\mu$ 是写成：

$$\frac{300x^2+1725}{x^4-40x^2+84}$$

 30

当然，从我们现代的观点来看，丢番图的记法看起来……不是这么难理解，但的确很难用来进行代数运算。它是一种笨拙的记法，即使丢番图告诉我们："在你彻底熟悉它之前，你会觉得它很难。"[5]因为没有代表加的记号，它必须把所有负项聚集在一起，放在代表减的记号后面。再者，他的记法没有让我们的思维意识到，x与x^2是同样的数类。

我们或许会说丢番图的记法极为笨拙，相较于我们今日的记法，它其实很难操作，但在那种情况下，他竟然可以做任何数学，着实令人惊奇。我们可能认为这样的记法必定阻碍了清晰的代数思维。也许吧，但惯用和熟悉会促进观念的发展。对我们而言，问题在于他的记法是以相同的外显形式表示所有对象，并且没有确实地区别运算符号（例如乘幂或求和）与数字或不确定的对象。乘幂之间若没有间隔——即利用加号和减号——我们的思维理解代数可能会更加艰难。[6]

---

[1] 我们在《算术》第五卷问题2观察到这一点：在一个几何数列中找三个数，使得每一个数加上一个已知数时，变成平方数。丢番图选择这个已知数为20，必须解$4x+20=4$。于是他说这是荒谬的（α τ ο π ο ν），因为等式右边的4应该是比20大的数才对。

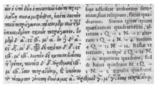
[2] 注意这些缩写也是第一个音节。

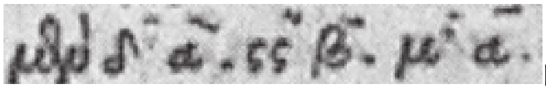
[3] 我建议将ars rei et census译为"'那物'的技术和性质"。我猜想所谓"那物"（rei），他必定是指"未知数"。

[4] 在巴歇的译本中，字母是小写，而且x的平方是一个奇怪的符号，但在其他译本中，符号是大写。

[5] 这是詹姆斯·高的翻译，参见他的《希腊数学史》附录， 108页。希斯提供了一个更正确的翻译，参见他的《亚历山大的丢番图》， 129页："或许这门学科看起来会很难，因为尚未熟悉它之故（初学者通常太早放弃希望）；但是你，带着你的热情之力和我的教学之助，会发现它很容易精通，因为渴望学习，当辅以教导，确保迅速进步。"

[6] 我们也欣赏用逗号将大数拆解成三个一组的方式。这样的设计可见于斐波那契的《计算书》，虽然他不是使用逗号，而是用近于括号的东西来将数字分组。



请注意在巴歇的译本当中，符号都是小写。巴歇谈到的1Q.+2N.+1是的译文。

第11章 负数是如何诞生的

代数的技术可能源自希腊人或印度人。然而，早在阿拉伯人学习代数、发展代数，并在11世纪末将这门技术带到西班牙之前很久，北印度的婆罗米人已经有了某种代数的概念。印度数学家婆罗门笈多撰写了《婆罗门修正历书》，收录一千零八首韵律诗，“供优秀数学家和天文学家娱乐之用”。该书完成于628年，它不仅提升了零在数学中的角色，也介绍了负数与正数的运算法则、平方根的计算方法，以及解各类线性方程和部分二次方程式的系统方法。

本书第7章谈到的10世纪著作《黄金草原和珠玑宝藏》提到一本更早的科学及天文学著作，名为*Sindhind*（梵语Siddhānta，音译“悉檀多”），这本历书中有记录太阳、月亮和已知行星的位置的天文表，以及星象数据和三角记号表。它是一本百科全书式的著作，记录了印度人所知的关于算术、天文学和所有其他科学的一切知识。

阿尔-花拉子密读完了《婆罗门修正历书》，很快将精力转向撰写阿拉伯文版的天文学专论《信德及印度天文表》（*Zij al-Sindhind*），该书是根据印度的信德人和印度人的方法写成，完成于825年之前。在研究特殊数学问题的过程中，他逐渐着迷于那些原本表示成文字形式及一部分为缩写形式的缺失量的求解方法。五年后，阿尔-花拉子密发表名为*Al-Kitab al-mukhtasar fi hi sab al-gabr wa’l-muqabala*的著作，概略翻译为《有关还原与对消的科学》或《还原与对消计算概要》，简称为《代数》。请注意阿拉伯文书名中的*al-jabr*这个词，它有时译为restoration或completion。事实上，这个词来自阿拉伯文中的动词to set（接合），如to set a bone（接骨）的用法。就像早期15世纪中叶古腾堡活版印刷发明之前许多其他手写稿的命运一样，阿尔-花拉子密的《代数》仅存最早的一批复本，年代不早于14世纪。而且，除了断简残篇之外，只有三部完整的复本留存下来。

帕乔利的代数专著是最早印行的代数书，阿拉伯文书名为*Alghebra e Almucabala*，即《还原与比较》。他也把这本书称为《大术：俗称在还原与比较这门技术中的共同规则》。

另外有些作者根据其他阿拉伯文字，提出关于代数名称起源的看法。16世纪的法国数学家皮埃尔·德拉拉梅（又名彼得吕斯·拉米斯）在其著作《算术》（*Arithmétique*， 1555年）中，提出让人半信半疑的观点：代数这个名称是叙利亚语，指一位杰出男性的技术与理论。他进一步提到，某位博学多闻的不知名数学家，写了一本名为《平衡》（*Almucabala*）的书献给亚历山大大帝，该书讨论暧昧不明或神秘难解之事，后来称为*Aljabra*，即代数的学说。

神秘难解，是的。但是，暧昧不明？*Almucabala*，切斯特的罗伯特的阿尔-花拉子密著作拉丁文译本的这个书名，听起来确实让英语系读者有暧昧不明和神秘难解事物的感觉。或许这两个名称都适用于这门技术。对于一位学习这门技术的9世纪波斯学生来说很可能神秘难解的事，对今日的学生而言稀松平常。《平衡》中提到一个例子：

> 我将十分成两个部分，使得其中一部分与另一部分的乘积为二十一。

书中没有要我们求出那两个部分。相反，它进一步提出求得这个问题答案的方法（括号部分是我的解说）。

> 现在，我们令根［x］表示其中一个部分，我们将它乘上十减掉根［$10-x$］，这表示另一个部分。所得结果为十根减掉平方［$10x-x^2$］等于二十一。十根部分补上平方［x^2］，并将此平方［x^2］加入二十一。由此得到十根［$10x$］等于平方加上二十一［x^2+21］。取半根，即五，接着自乘，得到二十五。从这个数减去二十一，得到四。取其平方根得二，以半根减去此数，余为三，此即为其中一个部分。

今日的代数学生会学到这种称为“配方法”的方法，用来处理二次方程式$x^2-10x+21=0$。

纯符号运算过程如下：

$x^2-10x+21=0$

$x^2-10x=-21$（两边各减21）

$x^2-10x+25=-21+25$（两边各加中间项系数之半的平方）

$(x-5)^2=4$（注意左边是一个完全平方）

$(x-5)=\pm 2$（两边各开平方）

$x=3$和$x=7$（两边各加5）

今日，阿尔-花拉子密的方法没什么神秘难解可言。配方法可远溯至古希腊时代，当时这类问题是以纯几何的方式处理，并且必须基于公理来证明。我们在《平衡》中看不到公理，或许这就是让该书显得神秘难解的原因。上述过程中所用的法则，遵循了某种偏向平衡方程式和开方求根的构想，也遵循了我们无法轻易地用9世纪的逻辑来表示的某种内在逻辑。

是什么让阿尔-花拉子密的研究成果不同于丢番图？《平衡》一书中的内容，除了一些小幅度的缩写改良、一种表示零的观念和符号，以及印度数字之外，看起来就像《算术》一样几乎都是以文字来表述的。它没有引进任何新符号。相反，该书列出一长串问题，根据不同的类别来组织安排。阿尔-花拉子密在该书第一页开宗明义写道：

> 我认识到数之还原与对消包含了三类，也就是根、平方以及数……
>
> 在这三种形式中，任两者可能彼此相等，举例来说：
>
> 平方等于根，
>
> 平方等于数，以及
>
> 根等于数。

所谓根，他指的是未知数，我们现在称为x。所谓平方，他指的是未知数的平方，我们现在会记为x^2。由此，用我们现在的符号记法，他的范例译为：

$$ax^2=bx,$$

$$ax^2=c，且$$

$bx=n$，其中，a，b，c皆为正数。

阿尔-花拉子密为我们提供了求解一般线性方程式和二次方程式的方法，将任一方程式简化为这三种形式之一。他的方法仅涉及在方程式的两边同加某量，以合并同类项到方程式的其中一边。这就是代数的运算方式：还原与对消。（在阿尔-花拉子密的时代，方程式不像我们今日的做法用某种等号来分隔两边——左边和右边，但如同分边做法的还原与对消概念应已合理可行。）代数符号只是提供了一种方

式，更容易执行这种合并过程；但在数学领域中，符号的使用并没有让代数得以做超过文字所能做的事。一旦了解了这点，整个代数研究成为从单一有限形式的观点，来看待一类数量无限的问题。

这种挣脱束缚的灵感引出了令人惊奇的发现。

阿尔-花拉子密写道：

> 当我仔细思考人们做计算时通常需要的东西时，我发现那总是数字……两个平方的和加上其中之一根的十倍，得其和为四十八个迪拉姆，两个平方的量为何？

这里，我们会写成$2x^2+10x=48$。

> 你一开始要先将两个平方简化成一个，而你知道两个平方之一是二者之半。接着，将叙述中提到的每一个东西化为自身之半，而这会让问题等同于“一个平方与该平方的五个根等于二十四个迪拉姆”，或者说，一个平方加上其根的五倍等于二十四个迪拉姆，该平方的量为何？
>
> 现在，取根数之半，其半为二又二分之一。

我们会写成$\frac{5}{2}$。

> 将其自乘，所得结果为六又四分之一。将此数加上二十四，其和为三十个迪拉姆又四分之一。

我们会写成$6\frac{1}{4}+24$。

> 取其平方根，得五又二分之一。从此数减去根数之半二又二分之一，余为三。此即为该平方之根，该平方本身为九。

我们会写成$5\frac{1}{2}-2\frac{1}{2}=3$。

阿尔-花拉子密写道“取其平方根”。**这个根**，是单数！对他来说，$30\frac{1}{4}$只有一个平方根，亦即$5\frac{1}{2}$。这使得他只求得一个解，也就是正根$x=3$。我们现代的解法稍有不同。我们会用高中代数学生所称的“配方法”，如本章开头的做法：

取$2x^2+10x=48$。

每一项皆除以2，得$x^2+5x=24$。

等号两边各加上$\left(\frac{5}{2}\right)^2$，得$x^2+5x+\left(\frac{5}{2}\right)^2=24+\left(\frac{5}{2}\right)^2$。

化简上式，得$\left(x+\frac{5}{2}\right)^2=\frac{121}{4}$。

两边各取平方根，得$x+\frac{5}{2}=\pm\frac{11}{2}$。

而最后，两边各减$\frac{5}{2}$，得$x=3$和$x=-8$。

哇！这里的-8是怎么回事呢？

我们知道阿尔-花拉子密的解法受限于代数的运算技术，因为连接代数与几何的数学手法，当时尚未为人所知。

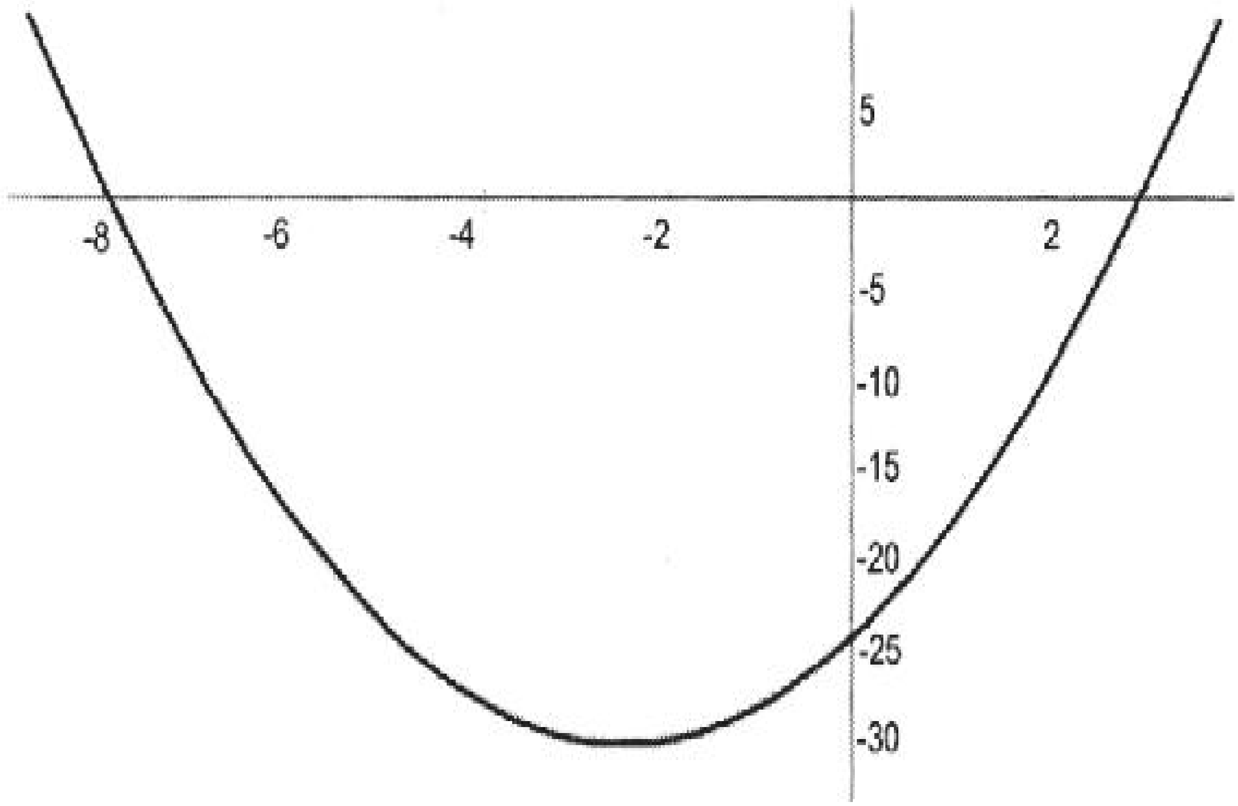


图11-1 $x^2+5x=24$的图形

利用二次方程式的几何思维，来推想阿尔-花拉子密的解。如果他知道并运用那种几何思维，他会将$x^2+5x=24$视为与高度为0的水平线交于两个x值——即3和-8——的图形（参见图11-1）。

或者说，如果他拥有符号代数工具，可分解$x^2+5x=24$，他会注意到他的方程式相当于（$x-3$）（$x+8$）$=0$，这个方程式有两个解，$x=3$和$x=-8$。

阿尔-花拉子密借由《婆罗门修正历书》来认识负数，甚至了解到任何形如$ax^2+b=cx$的方程式都有两个根。婆罗门笈多将零定义为一个数，一个数减去本身之后所得的数。依此，他可以列出他的算术法则：

一个负债减零等于一个负债。

一个财产减零等于一个财产。

零减零等于一个零。

零减一个负债等于一个财产。

零减一个财产等于一个负债。

零乘上一个负债或一个财产等于零。

零乘上零等于零。

根据他的描述，我们应可做出结论，就是他把负数当作数。他的算术逻辑法则谈到“财产”和“负债”，并由此衍生出关于正数的想法和负数概念的提示。他知道在某些情况下，一个二次方程式会有两个根，而应用问题中的条件会让那个方程式排除掉其中一个根。即使晚至斐波那契的时代，人们仍然怀疑负数根真的存在。

直到近期，婆罗门笈多仍被认为是在任何现代意义上第一位使用负数的数学家。但在中国，自第一个千禧年之始就已经使用负数。它们出现在《九章算术》中（参见第3章）。因此，中国人使用负数的时间，比印度人早了四百年。

第12章 数学史上的争斗

代数不是一直被称为algebra。15世纪中叶，一些意大利和拉丁文作者称它为“物与积之规则”（*Regula rei e census*）。数学家偏好用简短的名称来为他们的领域命名——算术、几何、微积分、分析、数论、逻辑等等。

韦达率先称它为“解析技术”（analytical art）。沃利斯给了它一个英文名称specious arithmetic（似真的算术）。他的这个命名最可能来自希腊文εἶδος，意指“类别”，并特别指未知数的**特殊**力量。“似真的”一词暗示一般用来表示所有已知量和未知量的类别——单子、平方、立方等等。15世纪时，specious这个英文词意指形式上看起来讨人喜欢但引人误解的。这个词在18世纪仍以这样的词义使用，英国辞典编纂家塞缪尔·约翰逊于1785年印行的《约翰逊词典》（*A Dictionary of the English Language*）中还可以找到这个词。牛顿称代数为“普遍算术”（universal arithmetic），可能是因为它蕴涵了用于求解一般方程式的所有普遍的算术法则。彼得吕斯·拉米斯认为那个阿拉伯文名称algebra，是“一门［如此］精妙的值得赞扬的技术”这个正式名称的通俗说法。对笛卡儿来说，代数的阿拉伯文名称是barbarous（未开化的）。

表面上看来，代数似乎是根据一些运算法则来进行符号运算的技术。但另一方面，我们该如何理解一个问题，它要求一个数，使其加上自身的五分之一后会得二十一？我们可以用语言将答案表示为十七又二分之一。这个特殊的问题出现在阿美斯纸草书中。公元前1550年之前写在那份纸草书中的解法，今日学代数的学生不会了解，他们会简单写成$x+\frac{1}{5}x=21$，然后用我们现代的代数法则轻而易举地使用那些符号，求解x，并得$x=17\frac{1}{2}$。学生们可能不会察觉到，有助于一般化、统合性及更深层理解的思维，全部直接来自记法本身。

20世纪数学家及科幻小说家埃里克·坦普尔·贝尔曾说——虽然他缺乏证据——在17世纪中叶，数学家之所以能够引进负指数和有理指数，是因为符号运算让他们的思考从杂乱无章的文字中解放出来。他在其广受欢迎的经典历史著作中写道：“那么，更多的荣耀要归于古人，他们坚持努力凭借错综复杂的文字来获取现代人用近乎机械化的几笔便能获得的成果。”

关于这一点，不乏证据，因为我们注意到，在许多世纪中，那些以文字表述形式来研究代数的数学家，无法看到我们今天所见的一切。利用新的符号来表示未知数的乘幂x，x^2，x^3……暗示着x乘幂的乘积如何被指数的加法法则支配。贝尔在他的著作《数学的发展》（*The Development of Mathematics*）中写道：“一堆难以置信的大量令人困惑的术语和无用的规则被扫入过去的历史中，随之带走了同样大量或更庞大的令人苦恼的想法。”

当今日的学生知道，在我们相当现代的负数观念出现之前，在我们可以用符号化的方式来表示负数之前，方程式$ax^2+bx=c$被视为与方程式$ax^2=bx+c$完全不同，往往相当惊讶。而上述这些方程式与$ax^2+bx+c=0$也不同。这或许显得奇怪，因为对我们来说，上述任一个方程式的解，同样也是其他方程式的解；然而，16世纪之前，任意的有理常数项a，b和c都必须为正。

根据我们对符号的定义，大体而言，除了精妙的数字和象形图案之外，巴比伦数学里没有其他符号。即使是丢番图，如同我们已经知道的，也没有运用符号运算的全部力量。但我们仍然运用代数这门技术，因为它过去不是——而且从来不是——仅局限于符号操作。代数——借由一种错综复杂的文字表述或借由符号本身——是理解关系的技术，其中“相等”只是众多关系当中的一种。

每个时代都会有令人惊叹的人物应运而生，不是推动人类的进步，就是让世界变得更美好——略举几位人物，如古腾堡、伽利略、达·芬奇、马丁·路德·金和曼德拉。我毫不怀疑像牛顿和莱布尼兹这样伟大的数学家，即使在一个没有任何符号的世界里，也能有了不起的数学成果。同样，我毫不怀疑在那样的世界里，他们的研究会变得极为艰难，而且遇到可能几乎无法克服的阻碍——**几乎**，因为对人类来说，几乎没有什么是无法克服的。所以是什么因素让符号和代数记法在16世纪下半叶大量涌现，那股迸发的力道强大到足以自17世纪后完全改变数学和科学的研究方式？

想象一下，如果代数仍然完全是以文字表述的，或甚至只是缩写，今日我们的世界会是什么样子？学代数的学生会害怕他们的初等代数教科书中通篇充斥的那些“文字题”。那些问题比15世纪的学生必须求解的问题简单，因为现代的学生知道他们要做的只是把问题改写成符号记法，然后再用符号运算的法则来想办法。

计算代数（abacus algebra）始自斐波那契的时代，14世纪中叶开始蓬勃发展。它是一种传统的解题方式，源自计算学校以及处理大数的算术和代数问题的计算师傅的论著，那些问题由运算法则或几何论证来支持。这项传统成为符号代数在16世纪上半叶开始兴起的部分助力。

是代数的概念促进了符号的发展，而非符号带来代数的发展。在雷科德注意到他可以设计一个符号 $=$ 来表示“is equal to”（是等于）这三个字，轻松“避免冗长的重复”之前，他已经在其著作《砺智石》中写了差不多两百次那三个字了。最初的诱因是需要让文字简略，但一旦相等的符号安排妥当，其他符号便沿用同样的方式处理。这个符号的简洁特色无意间带来一个好处：它在大脑里形成了一个简朴的图像，可以促进理解。

早期的史学家认为，最早用阿拉伯字母表中的字母来表示算术运算的阿拉伯人，是阿拉伯代数学家阿尔-卡拉沙第（al-Qalasādi）。他出生于巴斯塔（Bastah），位于今日西班牙东北部的摩尔人城市，在那里学习律法和古兰经。后来，卡斯提尔人开始东征，他往南移居到安达卢西亚的格拉纳达。15世纪早期，几乎整个西班牙和葡萄牙都被阿拉伯人占领，并且经常与卡斯提尔和亚拉冈（Aragón）的基督徒发生战争。阿尔-卡拉沙第写了七本算术书和一本代数书，书中以缩短的阿拉伯文和字母来构成数学记法。他的代数论著《算术科学的阐明》（*Al-Tabsira fi’lm al-hisab*）中的确使用了这样的记法。

他的记法显然是在尝试通过缩写的方式来让代数符号化，一种近似我们会认为是真正的符号的最早符号；然而，我们也应该注意到，北非的阿拉伯数学家已经使用这些符号至少一世纪，阿尔-卡拉沙第不是原创者。在阿尔-卡拉沙第之前一世纪，马格里布（Maghreb）[1]的数学家伊本·阿尔巴纳（Ibn al-Banna）和伊本·亚萨敏（Ibn al-

Yasamin）也已经设计出一种简略字母表记法的体系，而且可以确定的是，13世纪之前很久，东方就已经使用这样的字母符号了。

现在，小小的事情发生，再次推动人类智识的进展，并惠及全世界。想想伴随着自20世纪40年代以来便与我们同在的现代可程序化计算器革命而出现的语言工具，以及它们的快速变革如何为我们带来了现代笔记本电脑和网咖。仅依赖低阶机器或汇编语言，而没有通用计算机程序语言，我们能拥有那些笔记本电脑应用程序吗？当然可以！但我们得同情需要做那些工作的人，并且想象一下那得花多少时间。需要比受过教育的我们更具天赋的人才能忍受那样的单调沉闷和复杂程度。

所以，16世纪正是符号发展的早期阶段。没有16世纪晚期发展出来的符号语言，我们能拥有现代数学和物理学吗？当然可以！但要达到同样的成就需要付出什么代价？至少需要一大群人投身于数学领域。

阿拉伯人将代数这门技术带到西班牙后，意大利随即培育了这些漂泊到欧洲的代数种子。不幸的是，除了斐波那契的研究成果之外，我们对1300年之前的欧洲代数著作几乎一无所知。最早的那段时期的著作出自斐波那契、计算师傅保罗（Paolo de l’Abacco）和帕多瓦的贝尔蒙多（Belmondo de Padua）等人。到了15世纪末，代数的研究范畴尚未超出仅有一个未知数的二次方程式，程度接近今天的高中课程大纲。回顾当时，这门技术仍然是以文字表述的形式进行；那时还没有用来表示未知数或做运算的记号或符号，而且二次方程式只允许有正根。

1505年，费罗（Scipione de Floriano de Geri del Ferro，或更为人熟知的Scipio del Ferro， 约1465—1526年）解出一个复杂三次方程式特例，$x^3+ax=b$，其中a和b皆为正数。在当时，人们仍然怀疑负数的存在，因此没有使用负数。零也有同样的情况，尽管零已经传入欧洲超过四个世纪，人们依然认为零的概念很可疑。所以，方程式$x^3+ax=b$与$x^3=ax+b$被认为是不同的东西。在我们现代的符号代数中，我们使用负数和零，上述两个方程式与通例$x^3+ax^2+bx+c=0$没有太多差别。

费罗的解——就像当时大多数数学一样——不是用文字系数（literal coefficient）求得，而是策略性地选择方便的数字。解答

特殊三次多项式$x^3=9x+28$的公式解，并求得其解为$x=4$和$x=-2\pm i\sqrt{3}$，让我们有信心找到通用方法。人类的思维发展出自己的方式，通过反复出现的范例来处理抽象问题，但令人费解的是，它也从单一范例归纳出一般化规则。没有符号，肯定很难办到。求解过程中，可以清楚知道系数9和系数28可能是为了方便计算而设，虽然它可能局限了任何一般化程序的发展。这是费罗的时代处理代数的方式。

费罗所知的一般程序是，巧妙地利用变量代换，$y=x-\frac{a}{3}$，将一般的三次方程式化简为一个没有二次项的方程式，尽管它是以几何的方式进行，并且仅适用于特定的三次多项式，所以a代表的是一个特殊二次项的系数，一个特定的数。代换的想法，对于现代代数中将复杂问题简化的程序来说是如此普通，我们对这个才华洋溢、令人敬佩的构想必定感到惊叹，并且好奇在不使用符号的情况下怎么可能完成如此非凡的创造。这个概念不容易用文字表述，它必然借助符号化的巨大好处才能完成。所以费罗建构出三次方程式$x^3=ax+b$的一般解：

$$x=\sqrt[3]{\frac{b}{2}+\sqrt{\frac{b^2}{4}-\frac{a^3}{27}}}+\sqrt[3]{\frac{b}{2}-\sqrt{\frac{b^2}{4}-\frac{a^3}{27}}}$$

如果这看起来有点麻烦，只要想象一下，没有我们现代记法的符号，在费罗的时代，或甚至一百年之后，要处理这些问题是多么可怕痛苦的事。应用到$x^3=9x+28$，其中$a=9$且$b=28$，方程式整个右边“崩解”，得$x=4$。费罗不知道这个方程式应有三个解，但这点留待第14章再讨论。

到了1545年，意大利人——特别是卡丹诺、他的学生费拉里（Lodovico Ferrari）和他的对手塔尔塔利亚（Niccolò Fontana Tartaglia）——已经解出一般的三次方程式和四次方程式。同一年，卡丹诺的《大术》出版，这个书名是正式书名《大术，或代数规则，卷一》（*Artis Magnae, Sive de Regulis Algebraicis Liber*

Unus）的简称。它收录了截至当时所有关于三次方程式和四次方程式的知识，包括（首次印行）那些三次方程式有实数根和复数根（称为“真实的”和“虚构的”）。该书以几何的方式提出法则，并且承认是塔尔塔利亚提供给了卡丹诺解三次方程式法则——不是证明——的信息，而塔尔塔利亚是从费罗那里学到那些法则的。当时所有的代数主要仍是文字表述的代数，夹带着一些符号，例如以℞表示根。据推测，℞是代表“根”的拉丁文*radix*的简略写法；负数的平方根会写成℞ . $\tilde{m}$.，所以℞ . $\tilde{m}$.15代表$\sqrt{-15}$。费罗对一小群朋友、学生和同僚透露了他的想法，但从未发表。因此，我们没有费罗的著作作为直接证据证明这些概念源自费罗，唯一的说法来自卡丹诺，他曾从米兰前往波隆纳拜访费罗。

虽然他伟大著作的主要成就背后的故事，是数学史上最凶恶的优先权争斗之一，但卡丹诺的确推许过其他人，包括他的老友塔尔塔利亚。针对$x^3=ax+b$的解，卡丹诺在《大术》第一章承认：

> 在我们的时代，来自波隆纳的费罗解决了三次方与第一个乘幂等于常数之例，一项极为出色又值得赞扬的成就。由于这个技术超越所有人类能力的精妙，以及凡人天赋所能达致的简明，并且是一种真正的天赐礼物，也是对人类心智能力一种非常清楚的检验，任何致力于此道的人终将相信，没有什么是他无法理解的。为了效法他，我来自布雷西亚的朋友塔尔塔利亚，他的才智无法被忽略，在涉入一场与他［费罗］的弟子费奥（Antonio Maria Fior）的竞争时，解出了那个问题，而在我的深切恳求下，他将解法给了我。

他们之间的宿怨与上述承认无关，而是因为卡丹诺对塔尔塔利亚承诺将会保密，不会发表那项解法。卡丹诺提出他的版本的解法，一种纯几何式的解法，因为他没有解决这个难题的必要符号工具。举例来说，我们可以通过将（$a+b$）连乘三次，运用一些代数法则，来证明$(a+b)^3=a^3+3a^2b+3ab^2+b^3$。卡丹诺没有这样的优势，他必须以图解一个几何立方体的方式，来求得同样的结果。

他用文字表述的方程式有一个巨大的缺陷。它需要一长串的方程式，部分原因是他不使用零。他怀疑乘法的符号法则，在《大术》出

版后二十五年，他仍主张负乘负得正在数学上的真确性就像说正乘正得负一样。他写道："而故此，错误主要在于一般主张的负乘负得正，以免使得负乘负得正确实比正乘正得负来得正确。"所以很难理解他为何存有这种怀疑，却又支持线性方程式的负根并接受负数的平方根。

《大术》是代数中一部真正的突破之作。虽然书中没有那些即将发明出来的符号，也因此写得千辛万苦又读来累赘麻烦，但它让数学家有机会看到确实需要更好的符号，以便更容易更深化地理解。当时几何学更像是一门视觉逻辑的学科；为了促进内在逻辑，人们必须想象一幅画，至少是以心灵之眼观看。自毕达哥拉斯、欧几里得、阿波罗尼斯和阿基米德的时代以来，许多问题都像这样以几何的方法来解答——线、正方形，以及想象的空间立方体——用几何的方式证明，部分是因为没有太多其他方法可用。

举个例子，这里用纯几何推理证明代数恒等式$x^2-2ax+a^2\equiv(x-a)^2$。先考虑正方形ABCD的图形（参见图12-1）。令x表示边AD的长，a表示线段ED的长。则x^2是以AB为边的正方形的面积，ax是以ED和DC为边的长方形的面积，而a^2是以GF和FC为边的小正方形的面积。从大正方形面积x^2，减去两倍的ax面积，我们**几乎**可以得到以AE为边的正方形的面积。这里说几乎，是因为少了以GF为边的小正方形的面积（我们减了两次，应该只减一次）。所以，我们需要加回以GF为边的小正方形的面积。以代数的方式来说，我们的做法是先从x^2减去$2ax$，再加上a^2，可得$(x-a)^2$。因此，$x^2-2ax+a^2\equiv(x-a)^2$。

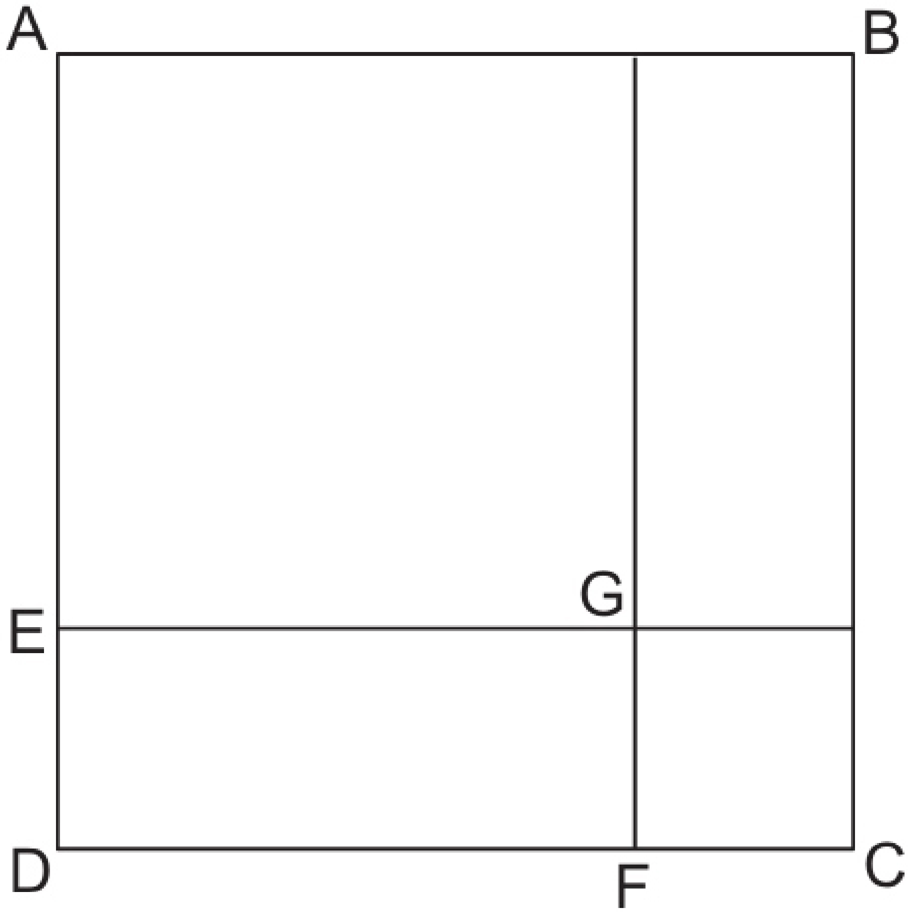


图12-1 以几何的方式检验$x^2-2ax+a^2\equiv(x-a)^2$

卡丹诺没有我们现代以实际数字来代表系数的明显优势。他在证明前述恒等式时，处理方式是选定特定大小的a值。所以他处理$ax^2=bx+c$这个形式的方程式时，会选定合于他的问题的正数a，b和c——例如1，10和144。在他的第一个例子中（《大术》，第五章），他告诉我们，$x^2=10x+144$的**那个**解是$x=18$（忽略了负根$x=-8$）。

若卡丹诺拥有必要的符号体系，可以直接进行代数运算，他可能会采取和我们一样的方式。根据法则来平衡方程式，他会将$(x-a)^2$看作$(x-a)(x-a)$，用分配律将右式化为$(x-a)x-(x-a)a$。接着，他会用交换律得$x(x-a)-a(x-a)$，然后再用分配律得$x^2+(-ax)+(-ax)+(-a)(-a)$。但到这里会遇到困难，因为他必须承认$(-a)$乘$(-a)$等于$+a^2$。他会反过来认为最后得到的$+a^2$不是因为$(-a)$乘$(-a)$等于$+a^2$，而是因为它是那块原本只需减一次，却被减了两次的面积。他甚至可能引述欧几里得《几何原本》

第二卷的命题7，以便确保我们可以用几何的方式来减和加正方形与长方形，由此得到想要的结果。即便他手边拥有适当的符号，他仍然需要有代数法则来验证他想进行的每一步运算，这样的法则在当时尚未完全准备好。

而谈到求解高于三次的多项式的详细证明，几何的帮助就不大了。“超出这一点就会变得愚昧了。”他写道，“自然法则不允许这么做。”所以《大术》难写又难懂，特别是谈到解高次方程式时。卡丹诺自己告诉我们，这是非常困难的工作。

卡丹诺的著作中最值得注意的部分是，虽然它是以文字表述的且处理起来很烦琐，但它明确认可虚数解和复数解，这些是早期作者完全避而不谈的。即使他没有将两个负数相乘以得到一个正数，但他对于乘和除两个平方根没有疑虑。

代数一直受到早期的语言和记法拖累，那些语言和记法阻碍了我们以一种符号化的方式思考。因此，我们很难确切知道我们的代数思维方式究竟起源于哪个时刻。想象一下，试着用未知量及完全以完整名称来表示的量做运算思考。人们痛恨烦人的重复；重复得太多的时候，他们开始寻求简化。

我忆起年方十五的那个夏天，在舅舅的网印工作坊（非法）工作了一星期。就像《摩登时代》里的卓别林一样，一整天从早上八点半到下午四点半，我像个单电路机器人般站在同一个位置，从网版上将墨还没干的海报移到干燥架上，一张接着一张：伸开双手，抓住海报两角，举高，转身，滑入架子，放开海报两角，转头，重复相同程序。无聊的时光里最有趣的部分是，一个架子满了，必须把另外一个架子推到指定位置时，会中断机器人般的程序，有机会走一两步。我计算着每一分钟，并且一张一张计算着移动了多少张没干的海报。让这个过程甚至更加难熬的，是一个就高挂在干燥架正前方墙上的巨大赛斯·托马斯（Seth Thomas）时钟，那个时钟很可能是为火车终点站设计的，它简直是世界上最慢的时钟。

每个晚上，我会发明和勾勒一个新奇的机械玩意儿，希望有一天它可以让白天那些单调乏味的人体传送带工作更轻松。可惜我没有精明地将我的想法申请专利……

人类总是制造工具或发明机械来做重复性的工作；他们对反复出现的问题，构思出抽象的解法；他们造出速记法来处理累人的冗长话语。也因此，写了数千次反复出现的算术文字后，数学家突然想到可以用文字的首字母来替换文字本身。

引用物理学家恩斯特·马赫的说法：

> 这或许听来奇怪，数学的力量在于它可以回避所有不必要的思考，并且在于它令人赞叹地节省了脑力劳动。即使是那些我们称为数字的排列记号（arrangement-sign），都是一种不可思议的简明又有效率的系统。当我们应用乘法表来进行多位数的乘法，从而使用已有的运算结果，而非每一次都重新进行整个运算时，当我们查阅对数表，已经进行过的计算因而取代并省去了新的计算时，当我们利用判别式取代总是要重新求解方程组时，当我们将新的积分式化为熟悉的旧积分式时，我们在这个简化的过程中，看见了一位拉格朗日（Lagrange）或一位柯西（Cauchy）的智识活动的微弱投射，他们借由伟大军事指挥官般的敏锐洞察力，以旧的已知操作取代新的操作过程。当我说最初等的数学与最高深的数学一样，都是有效率而有序的计算经验，准备好为人们使用，没有人会反驳我。

---

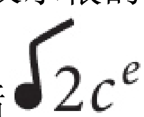
[1] 非洲西北部一个地区，阿拉伯语意为“日落之地”。古代原指阿特拉斯山脉至地中海海岸之间的地区，今日阿拉伯马格里布联盟包括摩洛哥、阿尔及利亚、突尼西亚、利比亚、毛里塔尼亚等国。——编者注

第13章 崭露头角的符号

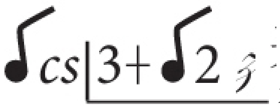
在卡丹诺的《大术》付梓之前一年的德国纽伦堡，人们正研读施蒂费尔的《整数算术》（*Arithmetica Integra*），一本关于算术和代数的专著。施蒂费尔在书中纳入了几种已在使用的符号，比如＋，－和√，而他也实际称这些符号为“加”“减”和“根”；然而，没有代表“等于”的记号。

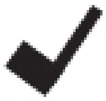
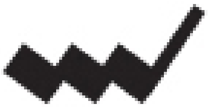
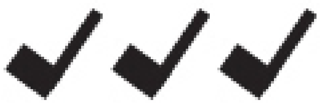
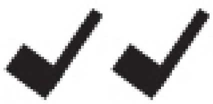
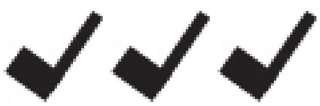
开始出现在欧洲代数手稿中的符号有两种相异的风格：一种来自意大利，另一种来自德国。意大利人使用cosa（“什么”或“东西”）这个词来指一个方程式的未知根。而由于代数基本上就是要找出这样的*cosa*的技术，欧洲北部的人开始称代数为Cossic Art（解未知数之术）。

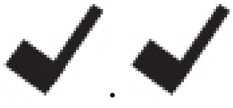
根数（平方根、立方根等等）最早的记法可追溯至约1480年，被开方数（被开平方的量或是被开立方的量）之前放一个点，就表示这是一个平方根：两个点代表四次方根，三个点代表立方根。到了1524年，这个点已经进化成一个带着一条往右上方弯曲尾巴的黑色圆点。

它看起来非常像一个音乐符号，像这样。还好，这个新符号很睿智地不是使用重复本身这个老方法来表示立方根和四次方根。相反，它使用另一个附加的符号，来表示根的等级（rank）。这样一来，表示的就是2的平方根，然后表示的是2的立方根。要表示四次方根的时候，会使用一个额外的点，因此，表示的是2的四次方根。

当时的代数只涉及了求解三次和更高次的多项式。以我们现代的记法，解答通常会以$\sqrt{x+\sqrt{b}}$这样的二项式来表示。一个拉长的L会和*cs*——拉丁文*communis*（共同的）的缩写——并用，来表示这个

根要被应用在整个二项式中。L的底部会在二项式下面延伸。例如，表示的是$\sqrt{3+\sqrt{2}}$，其中的加号在德国的手稿中已经偶尔使用。

鲁道夫的著作《求根术》（1525年）的施蒂费尔版本，就包含了代表平方根的符号（和我们的√非常相似），还有代表立方根的，以及代表四次方根的古怪。因此，64＝4，16＝2。这非常古怪，因为这些符号势必会造成64等于4，以及16等于2的各种误解。第二项是正确的，第一项则不然。很明显，鲁道夫并没有想在与的相似度之间建立任何关联，但史学家还是搞错了他的原意，将他代表立方根和四次方根的符号写成√√√和√√。这种混淆表明要设计简单易懂的符号，是多么困难的事。

鲁道夫所给的平方根符号，还有另一个不利的条件。他的读者如何分辨$\sqrt{\sqrt{512}+16}$和$\sqrt[4]{512}+\sqrt{16}$？他们必须仔细查看是否有那些点。为了说明他指的是$\sqrt{\sqrt{512}+16}$，他会写成512＋16，其中的点代表的是包含整个下一项的组。

《求根术》也许是以德文写成的第一本完整的代数著作。它是一本到16世纪早期所有已知代数知识的参考文献，更是未来教科书作者的绝佳资源。在它付梓的时候，大多数符号仍然只是文字的缩写而已，没有任何的标准形式存在。虽然＋和－偶尔会出现，但p和m也一

样常见。✔可以表示为R或℞，或者*res*（*x*）（拉丁文中的“东西”），代表的是必须找出来的未知事物。《求根术》出版后多年，史学家认为✔这个符号是字母“r”的速写。欧拉也这么认为。然而，如同我们所见，这个符号可能是从德文手稿里那个点发展成的符号，一种带条尾巴的点，也许是笔尖快速画下点之后所留下的痕迹。在施蒂费尔版本的鲁道夫《求根术》里，原本的符号✔并没有上面的横条，让这个符号看起来像是个字母“r”。

有个被称为比萨的达地师傅（Maestro Dardi di Pisa）的人，在1344年写了一份未出版的手稿*Aliabraa arbibra*。除了知道这个达地是算盘师傅，在将简略记法（abbreviated notation）引进中世纪数学中扮演了重要角色之外，这位作者的生平几乎无人知晓，而且这份被认为是最早以意大利方言写成的代数专门手稿中，甚至没有出现他的名字。*Aliabraa arbibra*里所使用的缩写，包括代表根的℞，$\widehat{\mathrm{m}}$（意大利文表示“较少”的词*meno*的第一个字母），代表未知数的*c*（*cosa*的第一个字母），以及代表四次方的*ce ce*（而不是*censo di censo*，平方的平方）。运算并没有被符号化；然而，书里包含了一些应该是用来教授乘法的有趣图表。达地的计算强迫他思考如何表示如$\sqrt{x+\sqrt{12}}$这样的平方根中的平方根，因为如果照以前的写法，会写成“℞ de zonto censo co ℞ de 12”。

鲁道夫所设计的平方根符号，被施蒂费尔于1553年稍微修改了一下之后，流传了下来。因此，到了1570年，德国的符号√进入欧洲，往西传入法国和英格兰，往南到了西班牙和意大利。

许多追溯回帕乔利的作者皆使用了℞这个记法，它很快变成了草写体的Ɍ，通常用来代表多项式的根。丘凯在他的代数著作《数学三篇》中使用了这个记法（以2为上标），但不是当作多项式的根来使用；他想用它来作为平方根的符号。我们并不清楚《数学三篇》到底是何时完成的，应该是1484年左右，但这份手稿是在18世纪70年代发现的，而因为几乎过了四百年才印行，它不可能对记法的历史发展有任何重大影响。当然，除了让丘凯的学生拉罗什（Estienne de la Roche）偷偷抄写了这部著作，并以自己的名义出版，获得撰写第一本法文代数书的荣耀。

对丘凯来说，一个数即是第一个根。如果他写的是Ɍ $^{1}9$，那他指的就是数字9。如果他写的是Ɍ $^{2}9$，那他指的就是3。1484年，丘凯写道：

> 一个数的根即是一个数，根据根的要求和本质，它自乘一次或多次，会精确地乘出根数所属的原数。或者相反，一个数的根满足：被写下或设定两次或多次，一个在另一个之下，或者一个在另一个旁边，然后，第一个乘以第二个，还有，乘上第三个，如有第三个，再乘上第四个，再乘下去，如有其他，最后的乘积将会等于这个数，或者将造出有那个根的数。而且我们应该知道有无穷多种根，因为某些是第二根，其他是第三根，再其他是第四根，还有其他是第五，继续下去，没有终止。

阅读那个时代的其他著作时，丘凯的记法Ɍ让人困惑，因为很不幸，它的祖先℞有两个意义。有时℞用来表示x^2，其他时候代表的则是x。这个混淆来自拉丁文*latus*（边）和*radix*（根），研究希腊文献的数学家使用这些写法，而在那些文献中，所有代数都被几何遮掩。如果我们现在考虑一个几何正方形，那么，℞代表的是它的

边x，还是它的边的根$\sqrt{x}$ √x呢？更让人混淆的是，这种记法有时指的是多项式的根，也就是对我们来说，会是所有未知数的可能值。唯一可以判断的方法，就是根据上下文来推断，所以读者必须非常细心。这个问题源于帕乔利，因为他将℞用在根和乘幂两者上。他在其《大全》（*Summa*，1494年）中写道：

$$℞.p^{a} = x^{0}$$

$$℞.2^{a} = x, \quad ℞.3^{a} = x^{2}, \quad ℞.4^{a} = x^{3}$$

以及

$$℞.2 = \sqrt{2}, \quad ℞.3 = \sqrt{3}, \quad ℞.4 = \sqrt{4}$$

如此笨拙的记法会引出问题。在上述的第二行中，那些数都比它们所代表的乘方要高一个单位。我们想用x^n来表示x乘以本身n次，这也是它应该拥有的意义，好让我们这个精妙的指数乘法公式得以成立：$x^m x^n = x^{m+n}$。帕乔利的记法有那个不需要的减1作为指数：

$$℞.m^{a} \cdot ℞.n^{a} = x^{m-1} \cdot x^{n-1} = x^{(m-1)+(n-1)} = x^{m+n-2} = ℞.(m+n-1)^{a}$$

我们可以忽略这种粗糙情况。这个公式是可行的，而且$℞\ .n^{a}$的确可以代表x的乘方，只不过不是n个复本。但是，粗糙并不是唯一的考虑。当时似乎没有一种可以用符号来表示乘方相乘的方式——也就是说，不回归以文字表述的方式就无法写出（x^m）^{n}，而我们也无法得到我们那个漂亮的公式（x^m）$^n = x^{m \cdot n}$。换句话说，如此笨拙的记法，正是更先进的符号表示法的一种阻碍。

虽然是个阻碍，但它还是有个优点。它将代数式的项，从它们古代几何的隐喻之中解放出来。自巴比伦时代以来，一个数的平方和立方就已经在思维里与几何的正方形和立方体联系在一起。欧几里得谈论正方形的时候，用的是δύναμις（“乘方”，也就是我们谈指数时用到的词）。当两条线的长度可以用同样的尺度（yardstick，用欧几里得的说法是pygon stick）衡量时，他会说两个量（magnitude）“在乘方上是可公度量的”。要使用同样的尺度来衡量正方形的对角线和正方形的边是不可能的。因此，对欧几里得来说，“乘方”纯粹是一个几何字眼，而不是一个由自乘所构成的数。

丘凯的数值记法℞^{1}，℞^{2}，℞^{3}，℞^{4}……已经超越了维度的几何比喻的可能性。一个直接的优势是，我们对这样一种观念如何延伸到负数会感到好奇。而丘凯的确对此感到好奇。他会用℞ $2^{1\overline{m}}$表示$\frac{2}{x}$的指数形式$2x^{-1}$。这种记法迫使x^0等于1，而丘凯知道这种关联并使用它。

倒过来想，《数学三篇》对现代指数记法的早期发展应该有所贡献，它会大大加速了代数的进步。但是事实上并没有。它比同时代数学先进了一百五十年；然而，由于它既未付梓也未被流传，因此数学家并不知道丘凯和他的《数学三篇》。也因为如此，负指数这个想法必须等到大约一百七十年后，当沃利斯的《普遍数理》使用负指数之后才浮现。

施蒂费尔在1553年出版了《求根术》的一个版本，我们的多项式[1]x^2-3x+2以1*Zm.* 3*Rp.* 2的形式出现在该书中。在这样的记法里，*Z*代表的是未知数，也就是x^2。它来自*zensus*这个词（拉丁文*census*的古德文拼法），意思是“……的数”（number of）。*R*代表的是*Z*的平方根，在我们现代的记法里，它表示的是$\sqrt{x^2}$，因此一定都是正数。这么一来，用更好的记法就可以揭示的负数根的想法受到阻碍，从而虚根有可能在一开始就被排除在外了。

到了1575年，该式子（$1Zm.3Rp.2$）变成了$1Q-3N+2$，后来变成$1AA-3A+2$。后面这个记法的优点是，它会清楚展示前两项的关系，也就是说，乍看之下，第一项和第二项除了乘方之外，几乎没有不同的地方。把这个多项式写成$1AA-3A+2$的时候，我们可以看出第一项和第二项的关系，但当这个多项式写成$1Q-3N+2$时，我们是“看”不出这个关系的。这个多项式的现代记法是x^2-3x+2。

---

[1] 此处原文误为方程式（equation），兹订正之。——译者注

第14章 笛卡儿的过人之处

“我觉得有必要说明这个在所有数学门类中，被今日大众称为代数的学科所拥有的最高地位。”邦贝利在1572年写道。邦贝利是工程师，工作和开垦沼泽地及建造桥梁有关。他的《代数学》（*L’Algebra*）于1579年出版，但早在二十年前，在托斯卡纳中部瓦尔迪基亚纳（Val di Chiana）沼泽的排水工程工作间隙时，他就已经开始撰写了。我们在《代数学》中见到了一种对于未知数及其乘方的新记法。举例来说，多项式x^2-3x+2的现代记法是从卡丹诺到笛卡儿时代历经中间几个阶段演化而来。以我们的标准来看，这些阶段之间的间隔既多又漫长。

邦贝利出版《代数学》之前二十七年，卡丹诺尚在撰写他的《大术》，这时等式的使用没有任何从属关系。拉丁文*aequalis*自由地用来表示两个式子是相同的，而且可以在没有损失的情况下互换。*aequalis*的两边有相等的地位（rank）。

几年后，雷科德在他的《砺智石》中将他的双子（Gemini）记号引进欧洲北部。它像是一个拉长的我们的“等于”记号，一个出现率比“＋”和“－”低，却比整个世界的海量数学文献里所有其他符号出现得更频繁的记号。

> 而为了避免“是等于”这些冗长的词语一直出现，我会如往常工作一样，设定一对并行线，也就是等长的双子线，因为再没有两个东西可以更能代表相同了。

这个双子，这个方程式[1]的至高符号，是一个送给我们现代符号集体的大礼。它的设计非常聪明（雷科德率先使用之前，也许意大利人也设计过），意在提醒有两个东西是完全一样的——一个有助于推理过程的简单发明。

邦贝利一定知道雷科德的著作《砺智石》，他想告诉我们一个式子引出另一个式子的时候，还是写了*fa*（“使得”）或是*eguali*（“等于”）这些字眼。在《代数学》中，你是不会找到等号的。我们在两个式子之间几乎不会看到*è eguale a*这些词。当我们希望有效说明两个东西是可以互换，并且在某种实际情形中没有差别的时候，

我们会使用“等于”这个词。我们可能会说四个25美分等于1美元。25美分硬币的成分主要是铜，1美元纸钞的成分则主要是木质纸浆和织物，以物理外观来看，它们是不一样的。有时候实际使用时，比如到便利店买口香糖，它们是没有分别的。但是，到时尚餐厅用餐，然后用25美分硬币付账，你就会知道差别在哪里了。

以意大利文写成的《代数学》在使用相等的时候，跟我们的观念是不一样的。诸如*fa*，*faro*，*eguali*和*eguale*这些词是单向的。说出*sommato 1 uia 1*，*farà 2*（“1与1的和为2”）与说出“1加1等于2”是不太一样的，因为在意大利文中，“2”是隶属于“1加1”之下的。在式子1＋1＝2之中有个概念上的不同，妨碍了等式两边的平衡。拉丁文*aequales*是“等于”的意思，代表的是两个可以完美互换的个体之间的无偏对偶性（unbiased duality），但邦贝利却选择了单向的*fara*。雷科德的记法非常适合且具优势的点在于，它所提供的正是当数学家需要两个个体在没有任何无心暗示的从属关系之下，可以自由互换的对偶性。

1572年时，邦贝利会将x^2-3x+2写成$1.\overset{2}{\smile}\ m.3.\overset{1}{\smile}\ p.2$，有时是$1.\overset{2}{\smile}m.3.\overset{1}{\smile}\ p.2\overset{0}{\smile}$，或是要将两个多项式相乘而把两行对齐的时候，写成$\overset{2}{\underset{1}{\smile}}.m.\overset{1}{\underset{3}{\smile}}.p.2$。我们可以在图14-1中看到他如何平方多项式$-4x^2+5x+2$，正确得到$16x^4-40x^3+9x^2+20x+4$。

让我们检视一下邦贝利如何执行这项乘法，看看他如何使用符号来做代数（参见图14-1）。在第一条水平线之下有两列，其中一列标记*più*（“更多”，意即“加”），另一列标记*meno*（“更少”，意即“减”）。在*più*列中，他将2相加得4，然后将*dignità* 1的系数相加得10，然后又是一个10。接下来，他将*dignità* 1的系数相乘得25，并进位到*dignità* 2。他再将*dignità* 2的系数相乘得16，并进位到*dignità* 4。这样*più*列中的所有计算就完成了。在*meno*列中也是用类似的乘法和加法。在不弄混*dignità*的情况下相加两者，得到正确答案如下：

$16\overset{4}{\smile}$.p.$9\overset{2}{\smile}$.p.$20\overset{1}{\smile}$.p.4.m.$40\overset{3}{\smile}$.


L'ALGEBRA
OPERA
Di RAFAEL BOMBELLI da Bologna
Diuisa in tre Libri.
Con la quale ciascuno da se potrà venire in perfetta cognitione della teorica dell'Arimetica.
Con vna Tauola copiosa delle materie, che in essa si contengono.
Posta hora in luce à beneficio delli studiosi di detta professione.
IN BOLOGNA,
Per Giouanni Rossi. MDLXXIX.
Con licenza de' Superiori.

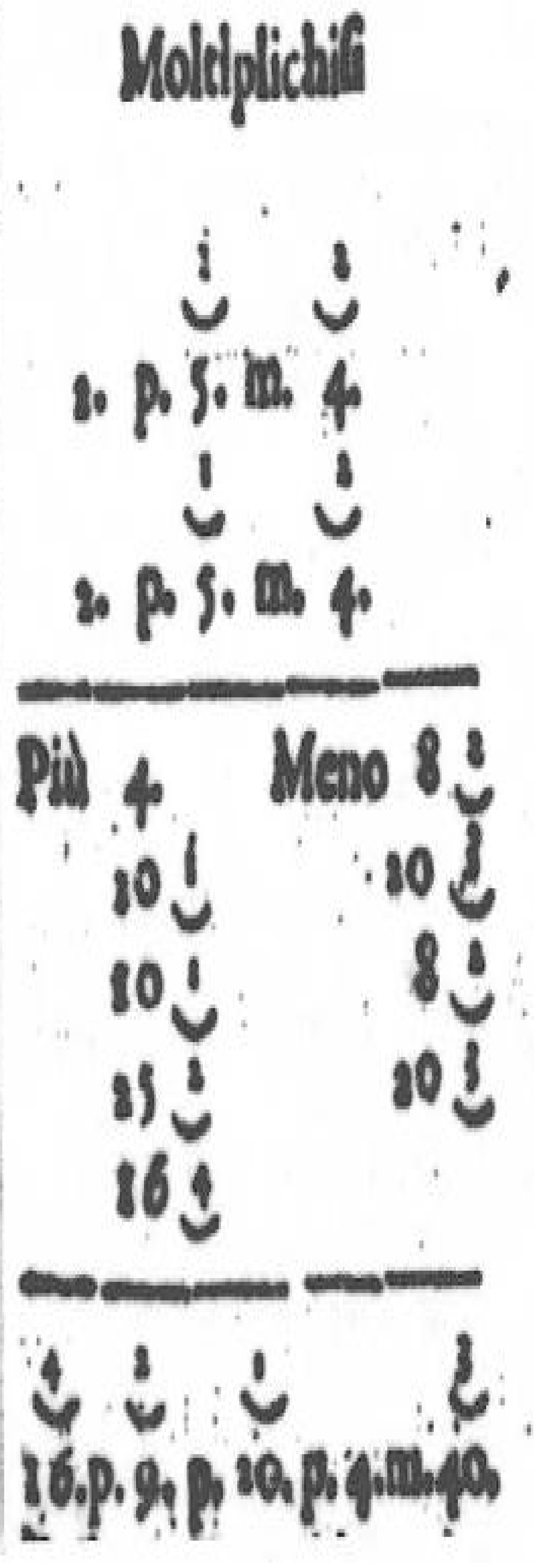


图14-1 《代数学》，封面和第二卷第217页，Centro di Ricerca Matematica Ennio De Giorgi

他所称的*dignità*就是我们现在说的exponents（指数）。现代意大利文中的*dignità*译成英文是dignity（尊严），对于我们称为“次数”或“指数”的东西来说，这个词看起来似乎有点奇怪。对邦贝利来说，更高的次数代表的是尊严的阶级制度。他的著作第二卷开头就是“Nomi delle dignità， e forma delle lore abbreviature”（尊严、值及缩写形式的名称）。然后他列出*dignità*如下：

Tanto $\underset{\smile}{1}$

Potenza $\underset{\smile}{2}$

Cubo $\underset{\smile}{3}$

Potenza di potenza $\underset{\smile}{4}$

Primo relato $\underset{\smile}{5}$

⋮ ⋮

Cubo di potenza di potenza $\underset{\smile}{12}$

我个人偏好的词是“尊严”而不是“指数”。exponent这个词来自拉丁文*ex-ponen*，我将它理解为“安置于上”，其中“于上”暗示了一种阶级，或是一种乘方的排名。古英语中的dignity意思是rank（地位）；莎士比亚用它来分辨两个相似但地位不同的东西：

然而，不同的肉身有不同的地位，

但一旦化为尘土，就再也没有分别了。

我们从这里可以看到邦贝利在将*dignità*画成托住数字的小杯子时，所创造的不只是真正的符号，也创造了新的数学名词。

这些精巧的、真正的符号将代数从几何学中独立了出来。邦贝利在手稿的前两百页里，使用了传统书写多项式的方式，也就是用Q（*quadrato*）代替Z，所以1*Zm.*3*Rp.*2会变成1*Q.m.*3*R.p.*2。

数的阶级制度让指数相乘的规则更显而易见。利用这些新符号，更容易看出：

$$1^{\underset{\smile}{n}} \text{ 乘 } 1^{\underset{\smile}{m}} \text{ 等于 } 1^{\underset{\smile}{n+m}}$$

那么邦贝利到底为什么给我们下列这么一长串*dignità*的乘积呢？

图14-2看起来像一群会让现代读者昏昏欲睡的羊的涂鸦。但最困难的事实是，我们容易从自己已知的观点去看待所有事情。我们知道$x^n x^m = x^{n+m}$，因为x^n在定义上不过就是n个x的乘积。因此，我们只需要数数x，就可以说服我们自己：n和m只要是任何正整数，就可以无限进行下去。然而，这么普遍的观念对于邦贝利所在世界的读者，可是非常陌生的。

邦贝利关心的是当a和b为正数时，三次方程式$x^3 = ax + b$的特定解。一个例子是我们在第12章提到的方程式$x^3 = 9x + 28$。一个解可以由简单的猜测$x = 4$获得，但另外两个解呢？这两个解会迫使我们做出荒谬的事情，比如取一个负数的平方根。

一般形式为$x^3 = ax + b$的三次方程式之解，是费罗带给我们的（参见第12章）：

$$x = \sqrt[3]{\frac{b}{2} + \sqrt{\frac{b^2}{4} - \frac{a^3}{27}}} + \sqrt[3]{\frac{b}{2} - \sqrt{\frac{b^2}{4} - \frac{a^3}{27}}}$$

1 uia 1 fa 2
1 uia 2 fa 3
1 uia 3 fa 4
1 uia 4 fa 5
1 uia 5 fa 6
1 uia 6 fa 7
1 uia 7 fa 8
1 uia 8 fa 9
1 uia 9 fa 10
1 uia 10 fa 11
1 uia 11 fa 12

2 uia 2 fa 4
2 uia 3 fa 5
2 uia 4 fa 6
2 uia 5 fa 7
2 uia 6 fa 8
2 uia 7 fa 9
2 uia 8 fa 10
2 uia 9 fa 11
2 uia 10 fa 12

3 uia 3 fa 6

3 uia 4 fa 7
3 uia 5 fa 8
3 uia 6 fa 9
3 uia 7 fa 10
3 uia 8 fa 11
3 uia 9 fa 12
4 uia 4 fa 8
4 uia 5 fa 9
4 uia 6 fa 10
4 uia 7 fa 11
4 uia 8 fa 12

5 uia 5 fa 10
5 uia 6 fa 11
5 uia 7 fa 12

6 uia 6 fa 12

图14-2 邦贝利冗长的*dignità* 乘积。资料来源：《代数学》，第二卷，第205—206页

将费罗的公式带入方程式$x^3=9x+28$，其中$a=9$，$b=28$。得到的解是$x=4$。那我们这些现代数学家知晓的另外两个解呢？表面上，费罗公式只给了我们一个解，即使立方根中的平方根不是负数。我们说过$x=4$，但如果我们实际将$a=9$和$b=28$带入，我们会发现$x=\sqrt[3]{27}+\sqrt[3]{1}$。的确，$\sqrt[3]{27}=3$，$\sqrt[3]{1}=1$。但如果我们真的想解出$\sqrt[3]{27}$，我们会设$x=\sqrt[3]{27}$，然后设法解出$x$。这也就表示我们要找出方程式$x^3=27$的解，而由我们的现代记法，我们知道$x^3-27=0$。等号的左边，会分解成（$x-3$）（$x^2+3x+9$）这个乘积。

这最后一个方程式有三个解，其中一个是由$x-3=0$得来的，其他两个则是由二次方程式$x^2+3x+9=0$得来的。当我们将所有解加入$x=\sqrt[3]{27}$和$x=\sqrt[3]{1}$时，我们得到$x=4$和$x=-2\pm\sqrt{-3}$，也就是$x^3=9x+28$的三个解。

想象一下邦贝利遭遇负数的平方根这些诡异的东西会怎样。在他的《代数学》第二卷中发生的事情非常奇妙，但如果我们继续探讨的话，会和我们原来的方向有所不同。如果想往该方向探索，请参阅巴瑞·马祖尔（Barry Mazur）精彩（也可以说“杰出”——如果这个词真的可以让大家不要怪罪我那真诚但可能偏袒亲人的溢美之词）的著作《数的想象》（*Imagining Numbers*）。而现在，我们可以满意地知道邦贝利遭遇并接受了虚数，甚至为它们创造了缩写记法。

più radice di meno（“更多的负根”）缩写为*più di meno*，并进一步缩写为*p. dm.*。因此，$\sqrt{-2}$会写成*p. dm.*2，而$\sqrt{-1}$只会

写成*p. dm.*。这时候，距离用符号i来代表$\sqrt{-1}$还有一段不短的时间，但*p. dm.*是个很大的进步，因为如果设计出适当的记法，计算上的错误会大大减少。欧拉在两百年之后书写的$\sqrt{-2}\sqrt{-3}=\sqrt{6}$和$\sqrt{-1}\sqrt{-4}=2$的错误，如果当时有记法能将第一个乘积记为$(i\sqrt{2})(i\sqrt{3})$而得$-\sqrt{6}$，第二个乘积记为$i(i\sqrt{4})$）得−2的话，这种情况是可以避免的。当然，我们必须原谅欧拉，毕竟这些错误发生在他的《代数原本》（*Elements of Algebra*）付梓时，而当时他已经失明了。虽然欧拉是将虚数$\sqrt{-1}$设为i这个符号的人，不过直到高斯于1801年使用之前，它并没有再度出现。

西蒙·史蒂文会将多项式x^2-3x+9写为$1②-3①+9⓪$。不知为何，那个*cosa*，那个“根”，那个“东西”——反正就是那个未知数的名字——是被理解的，因此在记法之中完全没有出现。1591年，韦达会将这同一个多项式写为*A quad* −3 *in A*+9 *plano*。然后，1631年，哈里奥特会将它写为$xx-3x+9$。他同时做了一件非常聪明的事。直到那个时候，数学家对于多项式的根，也就是会让多项式等于零的数，是很感兴趣的。哈里奥特这个巧妙的想法是，先将一个多项式设定为等于零，借此设定一个方程式，然后找出满足这个方程式的数。你可能会认为这只是一个语法上的差异，不会创造出任何新颖的东西，但是，这个想法却开启了一扇全新思考多项式的门。他看到多项式可以和一个数一样，由因子的乘积建立起来。举例来说，我们之后会看到$x^4-4x^3-19xx+106x-120=(x-2)(x-3)(x-4)(x+5)$。

这个巧妙的想法彻底改变了游戏规则：解多项式的根的问题，迅速地变成分解多项式的问题。一个简单的语法调整，促进了每个多项式方程式都有个实数根或复数根的可能性。不过，要用让人满意的严

密性证明它，可不是17世纪的方法能做到的。这样一个证明必须等待将近两百年，以我们现在所知晓的**代数基本定理**的一般形式出现，明确地告诉我们：任何一个多项式的根的数量（算上复数根）会和它最高的次数一样多。

更重要的是，这个想法赋予数学家一个恰当的代数形式的意义。将所有的项放在一个方程式（等式）的左边，然后让孤立的零放在右边，形式的重要性便会突显出来。这样，方程式$x^2+3x+2=0$的形式就明显会和$3x+2=0$不同，而$5x^3+2x^2+3x+2=0$又是另一种形式。代数不只关乎方程式，而且也和**形式**有关。

最后，笛卡儿于1637年在他的《几何学》中，想出了用数字上标来标记多项式的正整数指数。任何熟悉我们现代记法的人，都可以轻松地研读笛卡儿的这本著作。这个对我们现代人来说看似明显的简单想法，也就是将一个个乘方以数字等级排列，“次数”就是变量自乘的次数（换句话说，x^2就是$x\cdot x$，x^3就是$x\cdot x\cdot x$），立即改变了我们对多项式的看法和研究。笛卡儿的等级体系发展了从韦达和哈里奥特开始的，将未知数表示为x及将x^2表示为xx的书写传统。

---

[1] 原文是equation，本义是“等式”。——译者注

第15章 用声音来代表符号

处在一个由意大利人、德国人和英国人提供丰富数学贡献的时代，以拉丁名Franciscus Vieta写作的法国数学家韦达，研究时拥有非常大的优势。

他在其著作《引论》（*Isagoge*）命题2六个段落的一百三十九个词中，完整地将他对于π的著名计算展示出来。

Propositio II

Si eidem circulo inscribantur polygona ordinata in infinitum, & numerus laterum primi sit ad numerum laterum secundi subduplus, ad numerum vero laterum tertii subquadruplus, quarti suboctuplus, quinti subsexdecuplus, & ea de inceps continua ratione subdupla.

Erit polygonum primum ad tertium, sicut planum sub apotomis laterum polygoni primi & secundi ad quadratum à diametro.

Ad quartum vero, sicut solidum sub apotomis laterum primi secundi & tertii polygoni ad cubum à diametro.

Ad quintum, sicut plano-planum sub apotomis laterum primi secundi tertii & quarti ad quadrato-quadratum à diametro.

Ad sextum, sicut plano-folidum sub apotomis laterum primi secundi tertii quarti & quinti polygoni ad quadrato-cubum à diametro.

Ad septimum, sicut solido-solidum sub apotomis laterum primi secundi tertii quarti quinti & sexti polygoni ad cubo-cubum à diametro. Et co in infinitum continuo progressu.

他的命题2告诉我们如何得到π的近似值，他先将一个正方形内接在一个圆之中，然后将每边的等分投影到圆上，得到一个正八边形，接着一直重复这个步骤，从八边形开始，然后用获得的多边形继续。这是个阿基米德发明的老方法，不过，精简了一点，好让计算更简单。韦达以*Et eo in infinitum continuo progressu*（“然后我们持续增加直到无限”）这个六个词的句子，结束了他的命题。这是（就我所知）欧洲作者第一次使用无限重复一个代数过程的概念。在最后，我们以求解 $\frac{2}{\pi}$ 等于下面这个无限多项嵌套（nested terms）的无限乘积的方法，求出π：

$$\frac{2}{\pi} = \frac{\sqrt{2}}{2} \cdot \frac{\sqrt{2+\sqrt{2}}}{2} \cdot \frac{\sqrt{2+\sqrt{2+\sqrt{2}}}}{2} \cdots$$

对于这种嵌套平方根的无限和，即使是鲁道夫和丘凯也没有恰当的记法。不过，可以想象，在一些早期的德国手稿中，的确有可能出现这种记法。

韦达在早期可能会将多项式方程式$x^2-3x=2$写为*quadratum in A, minus A ter aequetur 2*，其中A代表的是我们会设定为x的未知数。其他时候，他会使用“＋”和“－”来作为代表加和减的符号，好让同一个方程式可以写成*quadratum in A, –A ter aequetur 2*。

他在之后写下了*X quadratum in A ter, minus A cubo, aequetur X quadrato in B*。这可以翻译成$3X2A-A3=X2B$（或者按我们的写法是$3a2x-x3=a2b$），也就是，试图三等分一个在弦为B、半径为X的圆中的圆周角之方程式。A代表的是该角未知的三分之一的弦。这里的变量是A，不是X。

韦达在这里展示了希腊几何与代数的密切联系，这是线、图形、立体的数学与符号代数的底层管道之间的联系，一种一直存在但几乎无人可完全理解的联系。的确，许多批注者非常清楚地看到了这些联系：亚历山大的希罗在1世纪想出了理解欧几里得几何的代数方法，而拉米斯则于1569年在他的《几何学二十七册》（*Twenty Seven Books* *of Geometry*）中写到几何与代数的关联。不过，让这个关联无比清晰的人，却是韦达。

在韦达的《几何学》（*Geometria*）中，他对于$3X2A-A3=X2B$这种形式的方程式非常感兴趣。它们就是我们现在所称的“齐次方程式”，也就是说，每一项中次数的和都会是3，而它的理念是只会将同维的项相加。欧几里得的《几何原本》中出现过这种形式的方程式。第二卷的内容都是长方形、正方形和其他图形的几何。如果你好奇几何学如何可以变成代数，你可以这样想：加或减这些运算，就像延伸或缩短直线一样；两个数a和b的乘积，就与一个有两个邻边a和b的长方形一样。求数a的平方根，相当于找出面积为a的正方形之边长[1]。我们可以从代数的观点看到这些关系，不过欧几里得要证明的是几何定理，而不是代数的。

代数的意涵可以从《几何原本》第二卷命题4来理解。该命题内容如下：

> If a straight line be cut at random, the square on the whole is equal to the squares on the segments and twice the rectangle contained by the segments.
>
> 如果一条直线随机地切成两段，直线上所张拓出来的正方形，会等于两个线段各自张拓出来的正方形之和，再加上由这两个线段所形成的长方形之两倍。

虽然这段文字翻译成了上述英文，但它的意义还是让人搞不清楚。*square on the whole*，*squares on the segments*和*rectangle contained by the segments*都需要解释。不过，请相信它们的意义在更详细的剖析之后，我们就会了解了。

将该直线的两个端点设为A和B，随机的切割点设为C。在这里，我们采用以字母——比如说A和B——来标记线段的惯例，将该直线标记为AB。另外，以字母标记四个角的长方形，会以四个字母并列表示。作一个边长为AB的正方形，将四个角如图15-1分别标记为A，B，E，D，然后画一条于C垂直于AB的直线CF。

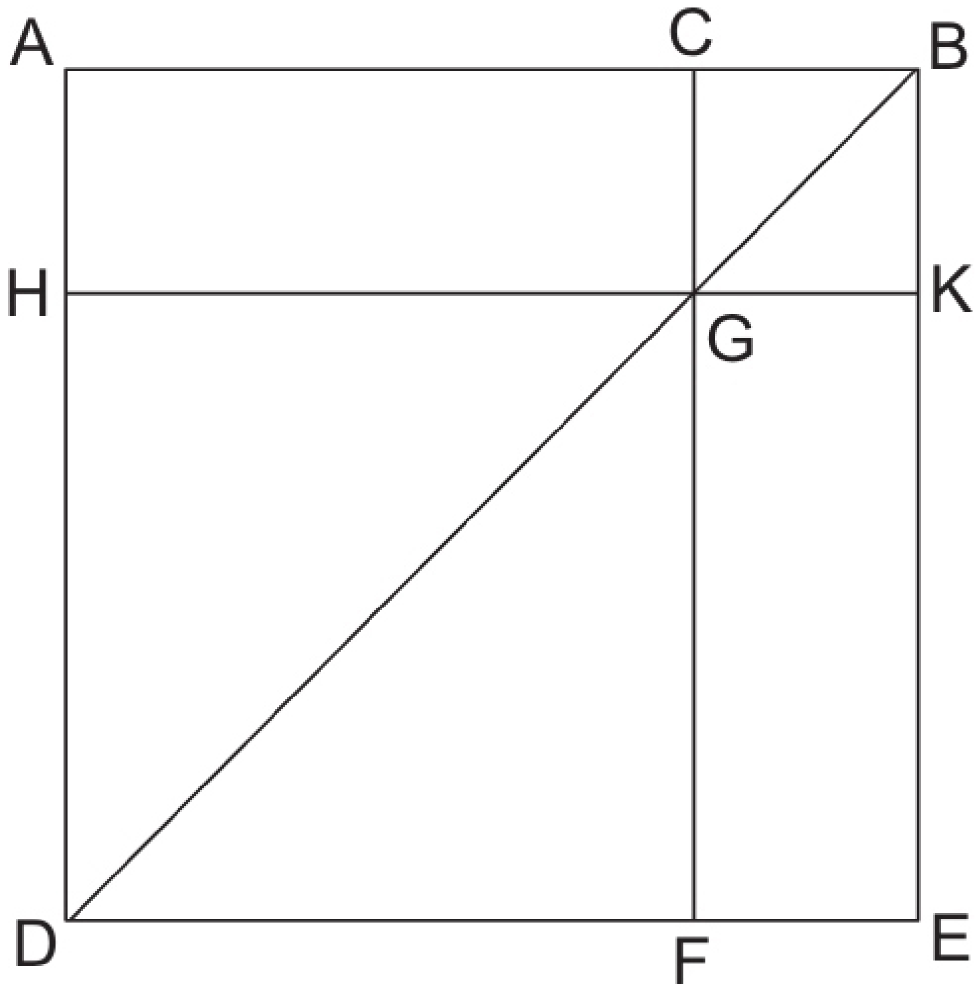


图15-1 整条直线张拓出来的正方形

现在，我们可以解释*square on the whole*是边长为整条（原始）直线*AB*的正方形，也就是正方形*ABED*（参见图15-1）。同理，*squares on the segments*代表的是边长为*AB*上的线段的正方形。这样的正方形共有两个，分别为*HGFD*和*CBKG*。我们将*rectangle contained by the segments*解读为一个边长为被*C*所切割的两个线段的长方形，也就是*ACGH*。从图中，我们可以看到*ABED*的面积等于*HGFD*的面积加上*CBKG*的面积加上*ACGH*的面积再加上*GKEF*的面积。

只要设$a=AC$，$b=CB$，然后用代数的方式来看那最后一句，我们就可以得到：

$$(a+b)^2=a^2+b^2+2ab$$

只要用一些基本算术法则，我们就可以证明这个等式。

不过，韦达在他的代数版本中仍然是使用文字和句子，而不是缩写和符号。尽管熟悉雷科德、邦贝利和西蒙·史蒂文的记法，看起来他似乎还是偏好使用文字。他对于代数的伟大贡献，并不是导入新的运算符号。韦达的著作中几乎没有新的运算符号，而是抽象地使用字母来表现涉及对象的更一般的特性，以及那些字母也可以和数字一样，遵从代数推导和法则的伟大想法。即使在阿尔-花拉子密的时代，人们也知晓方程式中的共同因子是可以消去的。韦达拓展了这个消去的观念，让我们知道如果$BE2+B2E=B3$，那么，$E2+BE=B2$，然后，如果$BE3+3B2E2=3E4$，那么，$BE+3B2=3E2$。换言之，非零的未知数也可以同样用已知数的消去法来消去。

之前为何没这样做呢？第一个答案是，如果多项式几乎全部用字母来表达，会让人非常困惑。我们将二次多项式写成ax^2+bx+c，可以立刻判断出哪些是已知数哪些是未知数，这个区别是早期的记法所无法提供的。第二个答案是，在抽象量上的算术运算（加、减、乘、除、开根号），看起来只是些没有什么计算优势的符号姿势（symbolic gesture）而已。写出3×4＝12，就是个将三个东西和四个东西聚在一起的轻松运算。但是，写成*B multiplicata per C aequalis B multiplicata per C*，到底能有什么优势呢？这就像是给出可能是赘余的同义反复$B\times C=BC$一样。

韦达使用元音来代表未知数，辅音来代表已知数。这样做有两个功用：它可以避免混淆已知数和未知数；更重要的是，允许我们使用一个以上的未知数。这样一来，一个方程式里面的未知数就可以区分了。举例来说，方程式$3X2E-U3=X2B$里面有两个未知数E和U，不过我们会将它写成$3a^2x-y^3=a^2b$，其中，a和b是已知常数，x和y是未知数。

韦达的字母系统还有另一个优点。元音A用来表示未知数量，连续的乘方则会标为*A quad.*，*A cubus.*，*A quad.quad.*。就像丘凯的系统一样，其中根数表示及排列为℞^1，℞^2，℞^3，℞^4……韦达的系统显示了一个单一未知数依次排列的乘方之间的关联，正如我们的记法x，x^2，x^3，x^4……虽然韦达的所有著作中都没有用到新

的运算符号，但他所导入的字母系统，对于数学的发展几乎是不可或缺的。

今日，我们可能会认为韦达的字母系统如此显而易见是很便利的记法，但甚至迟至16世纪结束之前，这样的观念仍是革命性的。为此，他有时也会被誉为代数学之父之一。

韦达的元音-辅音记法没有存在很久，但是它激发了符号代数极大的进展。对我们而言，似乎很难想象这样的想法有多聪明。对我们来说，字母代表固定的已知数和变量的未定数，是自然而然的。不过，我们是智识习性的产物。我们学习事物，并且忘记我们曾经有过必须学习以理解它们的一段时间。我们活在21世纪的第二个十年，现在可能发现手机和GPS卫星导航在技术上十分成熟，但在下一个世纪，当这样的现象被更先进的现象所取代，人们回顾并思考我们的进步时，想法就会如同我们现在看待比如说打字机或花园水管的简易现象。浸淫在我们自己常用的符号之中，的确很难想象韦达的想法，何以不曾出现在丢番图到邦贝利，这么多聪明的数学家身上。

我们可能会认为，在帕乔利的℞、斐波那契的*res*（东西）、丘凯的℞和韦达的元音-辅音记法之间，几乎没有概念性的差异。然而，是有差异的。韦达的元音既没有影响文化禁忌，也没有限制既存的何谓数之观念。℞与*res*这两者都是指它们所说的意思。即使是丘凯的℞，也是指它所说的意思。在任何真正的意义上，它们都不是符号，因为它们蕴含着事先形成的概念，代表它们所类似的事物。韦达给了我们某种比单纯的新记法更多的东西。他的*A*（我们的*x*）是一个真正的符号；它超越了它被假定去代表的对象之具体性（concreteness）。托比亚斯·丹齐克曾经说过：“正是由于这种转换的力量，**将代数提升到比便利的速记更高的层次。**”

然而，元音-辅音记法还有另外的优势。当我们将笨拙的字面表达式（literal expression）转换成更便利的等价形式时，我们可在韦

达的元音和辅音上执行运算。正是这种转换的优势，丹齐克再一次告诉我们，“将代数提升到比便利的速记更高的层次”。

还有另一项特点：提升代数的元音-辅音记法。想象一下代数将会像什么样子，要是不用一般表达式ax^2+bx+c，我们就必须详述系数a，b和c具体代表什么数。那意味着比如二次多项式x^2+2x+3之任意解的问题，将会被视为不同于二次多项式$2x^2+3x+1$，即使第一个多项式会快速地提出一个具体的程序来解第二个。每一个表达式将必须按不同的方式来处理，虽然一定会有解题的线索存在。韦达令人赞叹的元音-辅音记法给了我们一种思考与操作集体的、一般的、任意的（*the any*）、全部的（*the all*）的问题的方法。有关这一点，丹齐克也将会同意，它“将代数提升到比便利的速记更高的层次”。

更有意义的是它在形成广义的数字概念时所扮演的角色。在韦达之前，代数学家会将$x^2+2x=3$，$x^2-2x=2$和$x^2-2x+2=0$（借用我们的x代表未知数的记法）视为不同类型的二次方程式。在某种意义上，它们当然是不同的，也就是假定“平方与某物乘以一个未知数等于一个数”这种形式的表达，不同于“平方等于某物乘以一个未知数与一个数”这种形式的表达。这完全不是由于怯于使用符号，而是由于负数和零所引起的麻烦与两难。我们看到上述三个二次方程式同时具备了相同的形式：$x^2+bx+c=0$。

上述第一个方程式被$x=1$所满足，第二个被……嗯，它看来不像是具有有理根，而第三个也没有。我们现在知道第二个方程式有两个解：$x=1+\sqrt{3}$和$x=1-\sqrt{3}$。当我们试着求$x^2-2x+2=0$的解，最后得到的解之符号表示就像$x=1+\sqrt{-1}$和$x=1-\sqrt{-1}$。

最后这两个解在韦达的时代没有意义。不过，当方程式以一般项的形式，并以记法来表示成如$x^2+bx+c=0$时，我们发现其解恒为：

$$x=\frac{-b\pm\sqrt{b^2-4c}}{2}$$

这个一般解需要分类区别。

1.满足b^2-4c为完全平方之条件的那些可能解（solution candidate）：完全可接受的解。

2.满足b^2-4c非完全平方但$b^2>4c$之条件的那些可能解：可疑的解。尚未被接受为有效，尽管处于被接受的边缘。

3.满足$b^2<4c$之条件的那些可能解：完全没有意义的解，**复数**，其作为数的资格被否决。

德·摩根对这个问题的看法如下：

> ［韦达］归结道：减是一种缺陷，表达式里应该尽可能回避它。他的用词是vitium negationis（有缺陷的否定）。拒绝所有可能带来麻烦的事物，没有比这更舒服更省心的了。接下来，第二步……在于将代数结果视为必然为真，并且视为表现某种关系或其他，无论它们如何与导出它们的假设不兼容。一个特别的结果没有作为一个量的存在性一经证明，按照定义，就会被允许有另类的东西存在。针对这一点，我们没有进行特别的研究，因为那些法则——在其下可发现新符号将会给出真的结果——与那些先前应用到旧问题的法则，并无不同。……当抽象的负量之诠释显示，至少是困难的一部分允许有理数解，其余部分，亦即负量的平方根那些，被接受了，且其结果带着递增的信心，也被允许了。

从现代史学家的观点来看，德·摩根说得并没有那么正确。在韦达的时代之前很久，这些奇怪的类别就已为人所知了。毕达哥拉斯学派思考正方形和直角三角形时，已经遭遇了无理数，而卡丹诺在1545年的著作《大术》中对复数已略有考虑。然而，韦达的记法使这些“真实的”与“虚构的”根更明显了，因为这种一般性的记法揭露了一个重要事实：它们作为真实问题的居间解（intermediate solution），是有意义的——也就是，不管怎么说，这些代数解给出了正确答案，即使它们涉及没有意义的（解题）步骤。

它们在16世纪可能是没有意义的解，但在17世纪因为有了更一般性的记法，相较于从前，这种无意义引起了更多注意。因此引出了下面这个问题：何谓数？这是个基本问题，但这也是一个深刻的问题。它太深刻了，以至于需要它的时候才会被提出来。现在，我们有关数的更为成熟的概念，已经接纳负数的平方根成为家族的一分子。这种充实给了我们一个**代数基本定理**，即任意n次（$n\geqslant 1$）多项式恒有n个

可能同类或不同类的根。当然，那些根可能（且多半）是复数。然而，为什么说这个定理是基本定理呢？至少有两个原因：（1）因为它告诉我们每个多项式刚好是一次多项式之乘积，这种一次多项式的形式为（$x-r$），其中r为根；（2）因为它向我们保证：引出多项式的每个问题都有一个答案。

韦达更一般性的记法引起我们对另一个问题的注意：何谓形式（form）？方程式$ax+by+c=0$几乎都是字母，字母a，b，c代表已知数，而字母x和y则取未知数的全部值。我们视a，b，c为数值之表征，而不管它们具体是什么。因此，整个方程式首先且最要紧的是被视为x与y之间的关系式。不过，一旦这种关系式被建立起来，记法的这种奇妙的理解，允许我们进一步经由变化a，b，c的值（a，b，c是所谓的参数），来检视$ax+by+c=0$的形式，从而建立x与y之间的关系式家族（family）。一个方程式的形式因而成为一个新的研究对象，它引出了一类方程式，这在我们缺乏符号区别常数与变量这两组数值时，是不可能想象得到的。

韦达将其《引论》一书的最后四个词写成大写作结，并止笔于最后的句点。他写道：

> *Denique fastuosum problema problematum ars Analytice，… jure fibi adrogat， Quod est， NULLUM NON PROBLEMA SOLVERE.*
>
> 译文：最后，这个解析技术……正当地将问题中最最得意的问题据为己有，也就是，没有留下未解的问题（TO LEAVE NO PROBLEM UNSOLVED）。

---

[1] 原文未提及正方形之边长，兹订正之。——译者注

第16章 思维方式的抽象化

韦达去世刚好三十四年之后，笛卡儿出版《几何学》。该书提出有关记法的一种新想法，亦即一个法则：字母表前面的字母保留给固定的已知量，后面的字母（p之后）则用以代表可以取一系列数值的变量或未知量。笛卡儿似乎遵循了哈里奥特使用小写字母的方式，虽然他不承认曾经看过哈里奥特的著作。时至今日，字母表的这种以p为界的习惯，依然是宽松的标准法则。

德国哲学家丹尼尔·利普斯托皮斯（Daniel Lipstropius）是笛卡儿同时代的人且是其传记作者，他告诉我们：笛卡儿注视一只苍蝇沿着弯曲的路径爬行时，突然灵光一闪，得到他最卓越的想法。当然，这是一则传说，隐喻笛卡儿坐标系统源自他运用路径到墙壁的距离，来描述这一条路径，也隐喻一只苍蝇激发了数学上最激进的变革之一：代数与几何早期的合并。它是一则传说，因为笛卡儿坐标系统看起来一点都不像我们现代所用的那个样子，有水平轴与垂直轴来标示相关的变量。这个故事后来变得更离谱，说笛卡儿由于健康欠佳，每天早上都躺在床上，沉思所有的科学如何可以建造得像数学一样确定。

如果苍蝇在空间中沿着一条弯曲路径前进，它也将留下一条有算术数据（arithmetical data）的轨迹，而笛卡儿会理解这条曲线的几何可以由这些算术数据来重建；反之，这些算术数据也可以由曲线的几何来重建。几何与算术只是相同数学的不同诠释：代数与几何彼此亲密回响。美妙极了！

笛卡儿的确有早上赖到很晚起床而思考周遭环境及其个人存在的习惯。童年时期，他被允许赖在床上舒缓他无法控制的咳嗽，而到了下午，他的喉咙就好多了。对任何患有后鼻腔逆流的人，这是个不好的建议，因为早点起床症状才会纾解。然而，笛卡儿也许是借此思考物理世界，它是如何在本质上是机械性的，自然界的每一事物如何可以由力学定律来说明，所有的理论物理如何可以运用少数几个通则和自然界的可观察事实来表示，以及一小部分的原理和基本方程式如何可以运用代数方程式来表示。

那个神奇的勾股定理告诉我们：在直角三角形三边上的正方形存在一种结构性的关系。那个定理告诉我们一个求得空间中任意两点距离的简易方法，是怎么回事？我们有办法运用方程式和比例式，表示直线与圆锥曲线（椭圆、抛物线和双曲线，这些曲线都是以一个平面切割圆锥面而形成的），又是怎么回事？

几何与代数的亲密关系从柏拉图时代以来就已经备受怀疑，在那个时代，柏拉图学院的数学家探讨三等分角、倍立方体，以及化圆为方等标尺作图题。公元前3世纪，阿波罗尼斯研究由平面切割圆锥面所形成的曲线——椭圆、抛物线和双曲线。1世纪时，希罗想出一种代数方法，来计算曲面面积和立体体积。与丢番图同时代的几何学家帕普斯（Pappus）指出，几何与代数应该有某种关联。4世纪时，梅内克缪斯（Menaechmus）发现圆锥曲线与方程式的关系，还有，早期的希腊地理学家已经在自由地使用坐标系统。1361年，尼古拉·奥雷姆（Nicole Oresme）运用经纬线系统，引出坐标系的雏形，最后他利用水平线代表时间、垂直线代表速度。

几何学源自处理线、图形和立体的兴趣，这些对象是可以在头脑中想象的。代数则源自涉及数的问题——这些数是被数学中重要人物的几何概念所决定的。到了中世纪晚期，代数的发展集中到数的更抽象观念上，特别是韦达将记法推进到包括常量和未知量，这种记法将代数从几何隐喻的限制中解放出来。这可以朝向更一般的目的跃进，专注于抽象的几何量。

我们看到加、减、乘、除，以及开平方的代数运算，都呼应了几何学中的运算。不过，那些运算实际上真的可以操作吗？韦达知道两个数a和b的乘积，是等同于求作（标尺作图）一个邻边为a和b的长方形；而求a的平方根则等同于求作一个正方形，使其面积为a。然而，这实际是怎么做到的？

笛卡儿在他的《几何学》第2页中，教我们怎么做。首先，考虑乘法：假定我们有两条线段，以AB与AC表示。以任意方式或位置安排它们，但要把它们与端点A连接（参见图16-1）。

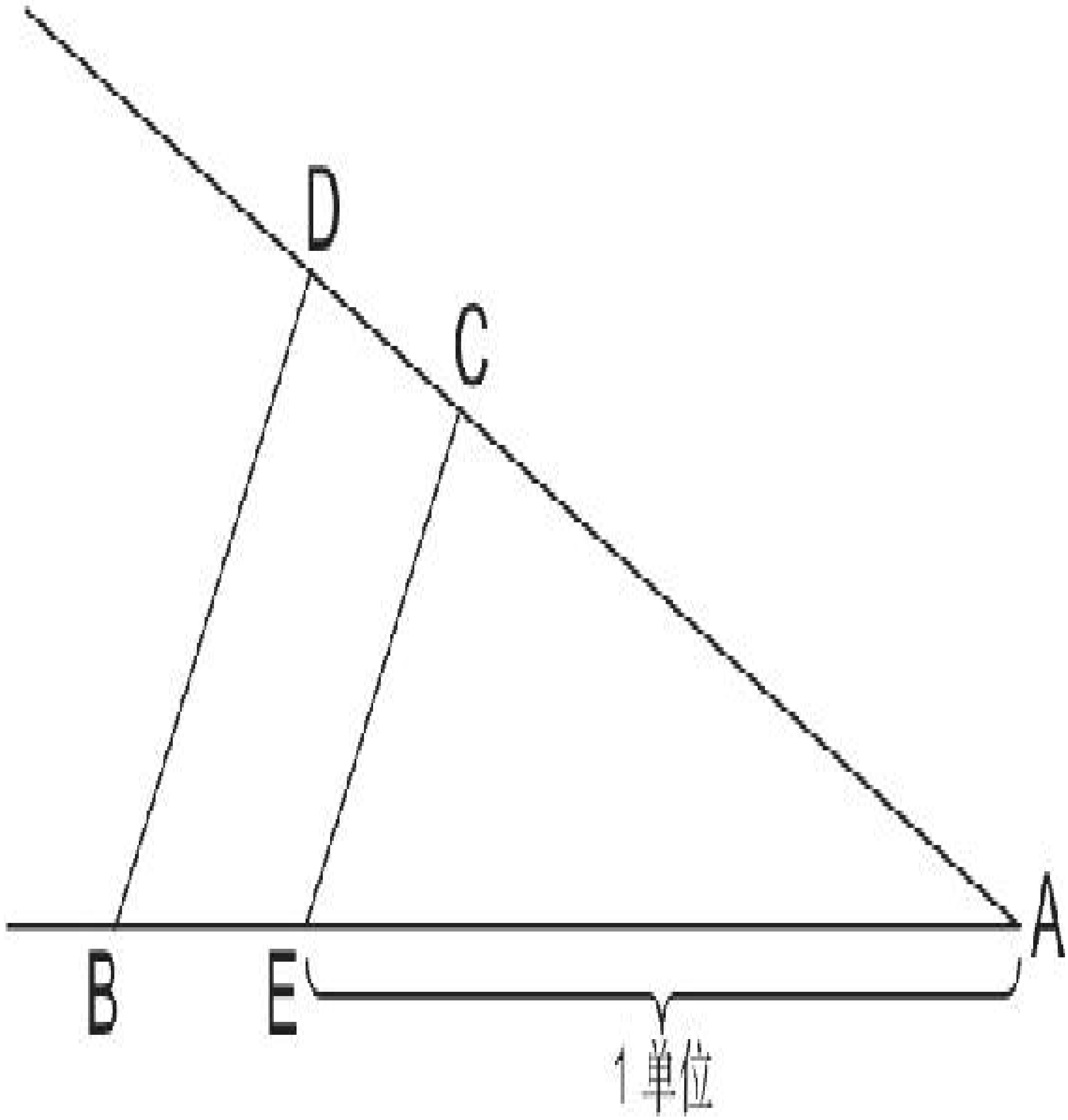


图16-1 乘法

在AB上，截出一个单位长的线段，并表示为AE。如果AB短于单位长，你可以延长AB。连接E与C，并作线段BD 平行于EC。由相似三角形，我们看到AD比AB相当于AC比1，因此，$AD=AB\times AC$。

针对除法，利用相同的设置；注意商AD/AB必须等于AC。要求AB的平方根，则延长AB一个单位到C点（参见图16-2）。

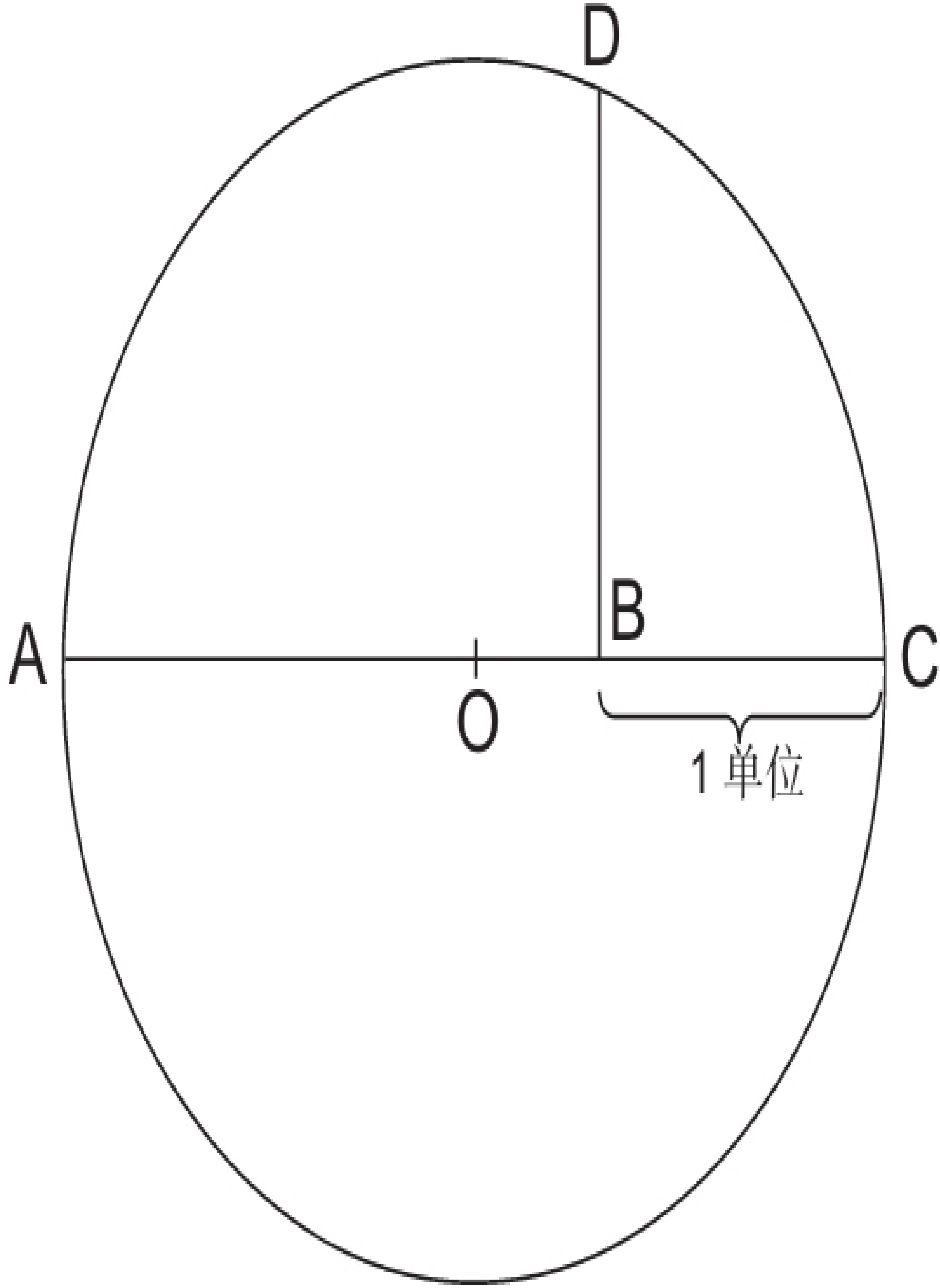

D
A
B
C
O
1 单位

图16-2 开平方

中分直线AC于O点。以O点为中心，作一个直径为AC的圆。过B点画AC的垂线，交此圆于D点，则$BD=\sqrt{AB}$。

所有这些运算都可以运用直尺和圆规来作图，因此，都可从欧氏公理来证明。而任意问题若可由标尺作图来表示，则也可运用一次或二次多项方程式来表示。

笛卡儿坐标系统不只是一个赋向的系统（orienting system），不只是从这儿到那儿的一条路径。它是一种通过代数透镜来看几何的方法。笛卡儿（费马也是）给了我们某种不可思议的奇特的东西。他向我们显示：思维本身有优选样式（optional mode）。他教导我们：为了将问题概念化，我们拥有优选样式。运用标尺作图方法，我们可能希望求解任意角是否可以三等分的问题，几何学的一个古老问题。这个问题可以自然地按几何方式，以文字如“线”和“角”来表达。然而，有时候我们会被我们认为自然的东西愚弄，被概念化的不必要扭曲所局限。

笛卡儿给了我们一种方法，以便在概念化的样式之间转换，将几何问题译解成为一个代数坐标系统。希腊几何学家的点、线和曲线，现在可以自由地运用抽象的代数表达式来表示，从而从我们有关空间的物理印象的束缚中释放出来，协助想象力在远远超出我们居住的可触摸世界之外驰骋，并进入一般性的非凡世界。

三等分任意角的问题，变成一个特殊的三次方程式是否具有一个有理根的问题。笛卡儿无法得知此一答案，不过，我们知道：问题中的根并不存在。

为了考察笛卡儿系统如何给予我们几何与代数的关联，我们要简短地提醒自己在中学数学课程中学到或漏掉的东西。请牢记：笛卡儿（以及费马）所发明的坐标系统，并不那么像我们今日所用的系统。事实上，它并不是最早的关于坐标系统的想法——14世纪的传教士奥雷姆有一个类似的想法。不过，笛卡儿的想法是今日更进步的概念（较晚引进）的一个启动器。一开始，我们观察带有一个随意挑选的固定标记的平面上的世界，就好比你可以选定一座高高的建筑，指引我们穿过一座陌生的城市。我们称此标记为（0， 0），这个符号的设

计想法到了下文就会显得明了（参见图16-3）。这个平面，就像计算机屏幕一样，有水平与垂直的数线通过（0， 0），它们的距离在箭头方向上为正，而在反方向则为负。

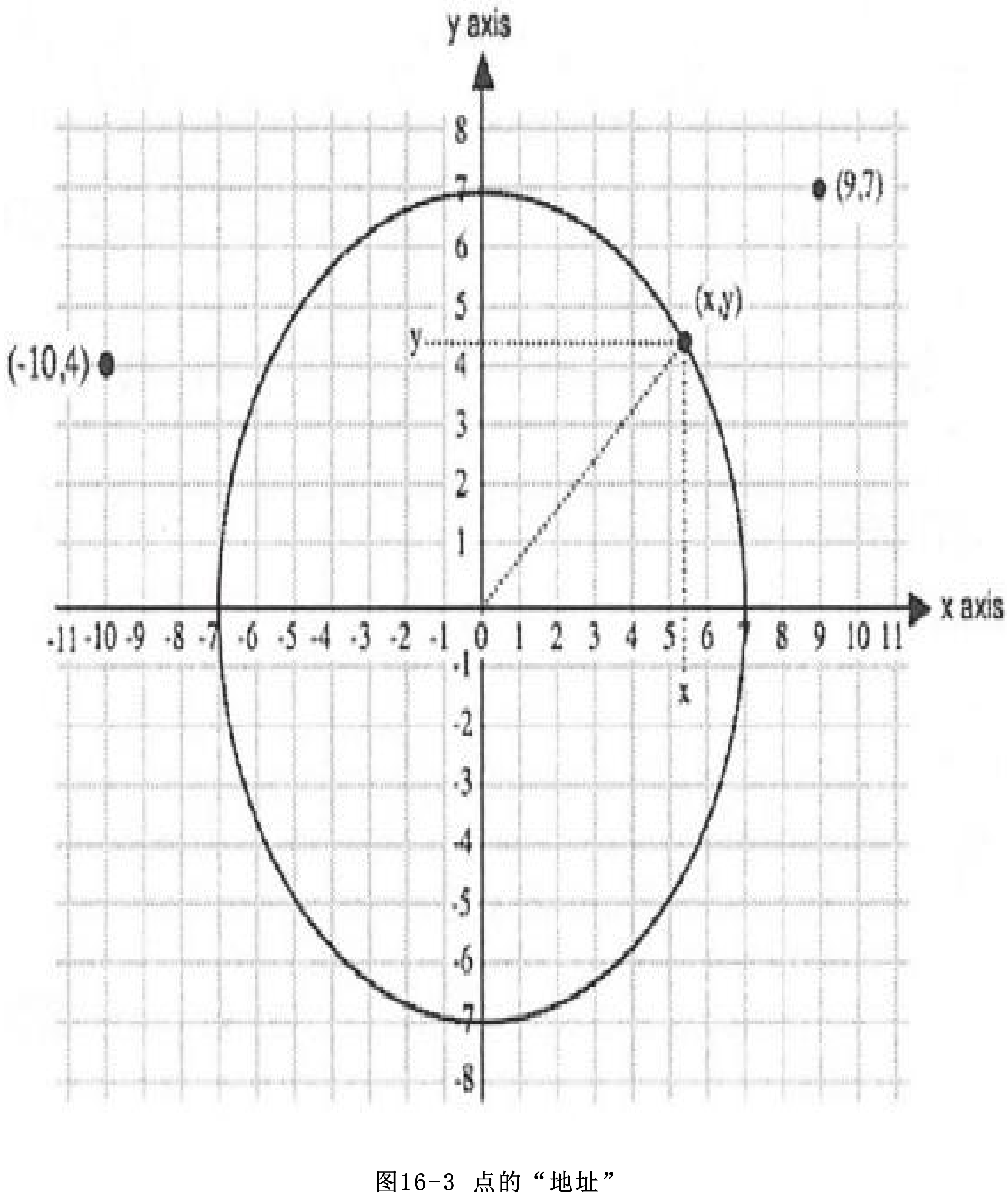


图16-3 点的“地址”

在这个平面上任取一点。暂时称它为P，那么，相对于标记点（0， 0）来说，P点位于何处？对P点位置的一个自然的描述（尽管还有其他的）是：相对于水平与垂直方向的箭头（使用任何你喜欢的单位），P点离（0，0）有多远（正或负）。举例来说，（9，7）是在右上角的黑点的“地址”，而（−10，4）则是在圆的左边的黑点的“地址”。

为了利用这个精妙的系统来表示中心在（0，0）且半径为7的圆，我们需要做的，无非是将圆上任意点（x，y）描述为过（0，0）的线段之端点，这个线段其实只是圆的半径，因此，在几何量上永远等于7个单位。对于这个底边为x、高为y的三角形，利用勾股定理，我们知道$x^2+y^2=7^2$。满足方程式$x^2+y^2=7^2$的每一对数值x和y，都将对那个半径为7的圆上的点，给出一个坐标地址。反过来说，该圆上的任意点将有一个坐标地址（x，y），其中x和y满足方程式$x^2+y^2=7^2$。我们用这个圆为例，只是想说明这种关联是如何出奇的简单。

看起来笛卡儿有信心使用他的解析几何的方程式，去解决几何问题。他解决这些问题，但总是想运用几何来确认他的代数证明。牛顿和莱布尼兹在他们的微积分中，也做了同样的事情。也许可以简单地说，他们都是全局型的数学家，总是希望看到全部头绪。

令人惊奇的是（除了使用小写字母的惯例），在使用字母表前面的字母来表示固定的已知量，用后面的字母来表示未知量之外，笛卡儿发明的新符号非常少。他改进了邦贝利和史蒂文表示未知量的指数的方法，使用上标来表示未知量的指数。哦——对了，还有就是——有一个与群组线（vinculum）有关的事。所谓群组线，是指与古老的德国平方根符号√相连的那条水平线，表示在这条线下的所有项都要放在一起作为一个整体开方。这就是我们现代的平方根符号 $\sqrt{\qquad\qquad}$。我们已经看到它是多么重要的一个进展。

笛卡儿在《几何学》中写道：

Et aa, seu a^2, ad multiplicandam a in se; Et a^3, ad eandum adhuc semel multiplicandam per a, atque ita in infinitum; Et $sqrt a^2 + b^2$, ad extrahendam radicem Quadratam ex $a^2 + b^2$; Et $\sqrt{C.a^3 - b^3 + abb}$, ad extrahendam radicem Cubicam ex $a^3 - b^3 + abb$, & sic de cæteris.

译文：而aa，或a^2，是将a自乘；而a^3，是再次乘a，等等，一直进行下去；而$\sqrt{a^2 + b^2}$表示求a^2+b^2的平方根；而$\sqrt{C.a^3 - b^3 + abb}$表示求a^3-b^3+abb的立方根，等等。

在此，我们有德国根号√与群组线的结合，用以涵盖一个求根的式子。我们现有的立方根符号$\sqrt[3]{}$还要再过三十年才会出现，它在几个地方同时现身：在米歇尔·罗尔（Michel Rolle）的《代数论著》（*Traité d' Algèbre*）中，以及莱布尼兹写给皮埃尔·伐里农（Pierre Varignon）的一封信中。

接着，我们在笛卡儿的《几何学》第4页中，发现多项式的书写方式与现在的写法颇为类似，除了笛卡儿使用了一个奇怪的符号∝来表示“是等于”。其中像z的图形不过是z的一个华丽的书写体。

$$Z \propto b,\ aut$$

$$Z^2 \propto -aZ + b^2,\ aut$$

$$Z^3 \propto +aZ^2 + b^2Z - c^3,\ aut$$

$$Z^4 \propto +aZ^3 + b^2Z^2 - c^3Z + d^4,\ \&c.$$

在第69页，我们首度找到方程式的完美可读的说明，看起来几乎就是出自20世纪教科书的形式。

Sciendum itaque, quòd icognita quatitas in qualibet Æquatione, tot diversas radices seu diversos vatlores habere profit, quot ipsa habet dimensiones. Nam si, exempli gratiâ, x supponatur æqualis 2, seu $x - 2$ æqualis nihilo; & rursus $x \infty 3$, seu $x - 3 \infty 0$; & multiplicetur $x - 2 \infty 0$ per $x - 3 \infty 0$; habebitur $xx - 5x + 6 \infty 0$, seu $xx \infty 5x - 6$. quæ Æquatio est, in qua quantitas x valet 2, & præterea etiam 3. Quòd si rursus fiat $x \infty 4$, atque $x - 4 \infty 0$ multiplicetur per $xx - 5x + 6 \infty 0$, producetur $x^3 - 9xx + 26x - 24 \infty 0$. quæ alia eft Æquatio, in qua x habens tres dimensiones, tres quoque habet valores, qui sunt 2, 3, & 4, atque una falsa, quæ es 5.

在上述引文中，笛卡儿写道：如果我们将多项式$x-2$乘以$x-3$，其结果将是x^2-5x+6。如果我们将后者乘以$x-4$，则得到$x^3-9xx+26x-24$，而如果我们继续做下去，将最后这个多项式乘以$x+5$，则可得$x^4-4x^3-19xx+106x-120$。因此，$x^4-4x^3-19xx+106x-120=0$的根就是2，3，4及−5，从而$x^4-4x^3-19xx+106x-120=0$与（$x-2$）（$x-3$）（$x-4$）（$x+5$）$=0$只是同一个方程式的两种不同表示方法。[1]

在前一章中，我们看到韦达如何引进字母表的辅音字母来推进代数的发展。在字母被用于那一类表示之前，丘凯、邦贝利和史蒂文有关多项式的记法完全够用。一个已知数的平方或立方可以计算而得到另一个已知数；当你只是表示4时，我们根本不需要写成2^2。在任意多项式中，只有未知量被提成乘幂，而未知量的乘幂（unknown-to-a-power）无法直接计算得知。因此，对16世纪的手稿来说，写3^3表示$3x^2$，3^1表示$3x$，以及3^0表示3，其实无伤大雅。丘凯、邦贝利和史蒂文运用这样一个所谓指数计划（index plan）来书写指数，不致有歧义。

然而，还是有个问题。一个多项式可能会有多于一个未知数，比如说一个x、一个y，如此一来，$3x^2+5y^2$在指数计划下就写不出来。我们偏爱笛卡儿的记法，主要是因为它是我们自己的记法，但也是因为它比之前出现的有歧义的记法来得讨喜，还因为没有人想出更好的符号……

一个符号可能变成常规而持续使用数世纪，直到某事发生产生窒碍难行的情况，而推动脉络的进展。譬如，16世纪的指数计划就出现了这样的情况。现在，多项式代数记法有一种合用的记法，在下一个千年应该不会改变。帕乔利的R_x断断续续使用了近两百年之久。鲁道夫的√相较于其他较无价值的平方根表示方法，有一些改进，不过在一百年间没有什么变化，直到笛卡儿加上一条群组线。

在笛卡儿的《几何学》中，我们看到自己使用的平方根符号（参见图16-4上部），在德国平方根符号√上加上了群组线，以便统合将要开方的所有项。某些人必定认为R_x会一直使用下去，就像我们认为我们的指数记法会一直使用下去一样。

$$Z \propto \frac{1}{2}a + \sqrt{\frac{1}{4}aa + bb}.$$

Quòd si verò habeatur $yy - ay + bb$, atque y sit quantitas, quam invenire oportet, facio rursus idem triangulum NLM, & à base ejus MN ausero NP, æqualem NL, eritque reliqua PM, æqualis y, radici quæsitæ. Ita ut fiat $Z \propto -\frac{1}{2}a + \sqrt{\frac{1}{4}aa + bb}$. Nec aliter sit, si proponatur $x^4 \propto -ax^2 + b^2$. PM enim esset x^2, & haberetur $x \propto \sqrt{-\frac{1}{2}a + \sqrt{\frac{1}{4}aa + bb}}$: atque ita de aliis.

图16-4 笛卡儿《几何学》（1659年）第3页中的群组线

在上述引文中，有一个平方根的嵌套出现。通过它，我们不需要太多想象，便可以预见可以利用一个连续嵌套显示韦达近似$\frac{2}{\pi}$的精彩命题2（参见第15章）。

数学史家弗洛里安·卡裘利告诉我们，是笛卡儿引进了这种新的根号、群组线等等。奇怪的是：是谁最先想到群组线的？凡司顿（Francisci van Schooten）在1648年编纂韦达的《数学全集》（*Opera mathematica*）时，已在他的批注中使用了群组线。嗯……图

16-4引自1659年出版的笛卡儿《几何学》的凡司顿版本。凡司顿有没有可能为了简化笛卡儿的意思，而将群组线夹带了进去？

对我们来说，幸运的是，身为数学家的笛卡儿有着伟大的影响力，而且他能够推动最佳记法的标准化，让它们走向下一个世纪。17世纪充斥着许多实验，当时所运用的各种怪异又累赘的记法，很有可能长期阻碍数学的进步。

对希腊几何学家而言，一条曲线或多或少是一个静态图形（static figure）。笛卡儿开始以不同的方式思考曲线。他的坐标系统被认为是由一个法则（即其方程式）所决定的一组动态位移（dynamically moving）的点之组合，亦即一种带着地址的代数对象，这些地址（亦即点）是由实数x和y来表示的。那些实数，也就是“坐标”（coordinate），被一起锁在一个共同有序的（co-ordered）数值关系式中；其中一个无法变化，除非它得到另一个的允许。这种新几何将曲线视为变量之间的关系。这是一种巨大的进步，大大改变了数学的策略和样态，让微积分的问世变得可能，并且永远改变了我们思考运动现象的方式。

17世纪早期一种典型的经验观察，会将抛射体在不同时间的高度显示为一个数值表。当抛射体未被观察的时候，就无从寻求它的高度。有了图形的观念，以及一个建立高度h与任意时间t的关联的代数方程式，针对抛射体高度如何随着时间的改变而平滑地改变的现象，我们的直观理解就会出现，那就是，数值如何爬升与下降的一个图像。

几何与代数的统一性是数学史上最伟大的发现之一。一次到位，它让我们对支配一个事件与多个相依事件之间的关系的定律，有了理解。它为后世的数学家提供了有力的工具，让他们得以针对两个相关的现象中，其中一个的变化如何影响了另一个的变化，按数学的方式来描绘与清晰表达。

然而，科学家对于大自然是否完全是机械式的，或完全可以用数学来说明，一直莫衷一是。现在，几何与代数的这种新的结盟指出，宇宙的秘密可以完全用数学的方式来说明。空间与时间联系在一起，不仅通过直觉所领会到的不确定、不可靠的几何图形，而且也运用了代数。

利用函数概念检视空间-时间关系式是自然而然的，但那首先必须等到1692年莱布尼兹引进一个原型概念（proto-concept），然后是约翰·伯努利（Johann Bernoulli）和欧拉做出某些改进，继而等到1834年狄利克雷引进他的版本。

---

[1] 此段原文有误，兹按上下文修订其方程式。——译者注

第17章 加减乘除的用法伊始

1660年6月13日，星期天，奥特雷德去世，享年八十八岁。英国博物学家约翰·奥布里（John Aubrey）告诉我们：“他个子矮小，一头黑发，黑色眼珠（炯炯有神）。他的脑袋瓜子一直动个不停。他会在尘土上画线和图形。他已经烧掉所有论文，宣称：‘这世界不值得拥有它们。’他是如此了不起。他甚至烧了多本印刷出版的著作，眼看着它们烧成灰烬，始终不动声色。”如果你端详捷克雕刻师温策尔·霍拉（Wenzel Hollar）为他雕刻的肖像，会发现一个上了年纪的薄唇男人，有一个从眉毛顶端延伸下来的大鼻子（参见图17-1）。

GULIELMUS OUGHTRED ANGLUS, &c.
ex Academia Cantabrigiensi, Aº ætat: 73: 1646.

W. Hollar ad vivum delin: 1644 — fecit Antuerpiæ Aº 1646.

图17-1 奥特雷德

他在1631年完成他的《数学之钥》第一版。后来有许多版问世，且在他死后半个世纪间，该书是广受欢迎的教学用书。在该书中，我们发现圣安得烈十字（St. Andrews cross）“×”首度用作乘的符号。这个符号在中世纪时期即已用来表示无关两数相乘的许多事物。一直到奥特雷德的《数学之钥》问世，乘法才被表示为并列形式，ab亦即a乘b。这没什么不好，只要被乘的本身都是符号。不过，当涉及确定的数时，那就有了歧义。譬如，22是表示二十二，还是2乘2？

并列的用法并非符号；它是一个会引出混淆的记法的概念。1545年，施蒂费尔分别使用字母M和D来表示乘与除。1585年，史蒂文也是如此。他们会写3②D sec ①M ter②来表示：

$$\frac{3x^2 z^2}{y}$$

其中，sec代表“第二个未知数”，而ter代表“第三个未知数”。又一次，我们看到M，D，sec和ter都不是真正的符号，只是缩写而已。它们很容易产生概念上的混淆：哪个未知数是第一个？哪个是第二个？或哪个是第三个？由于笛卡儿的贡献，我们现在的记法避开了此一问题，因为字母表内的字母早已按顺序排列了。

韦达书写“A in B”来表示A与B的乘积。迟至20世纪初，有些作者还在使用M来表示乘。甚至今日，还是有一些歧义与（乘法的）“并列”有关：我们写$3\frac{1}{2}$来表示$3+\frac{1}{2}$。或许这就是如此多年轻学生在计算带分数时，会犯那么多错的原因吧。

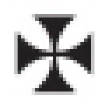
在奥特雷德引进的一百多个具有前瞻性的符号和卷标中，只有不到十二个还在使用。不过，任何人能设计出六个足以成为标准的符

号，存留三个世纪以上，都是值得喝彩的。

到了17世纪，大部分以文字表述的数学文献都转换成符号形式。各式各样的新记法引进，有些方便使用，有些则否，有些不实用，有些则是够蠢了。然而，前进的脚步并未停歇。埃里冈在他出版于1634年的著作《数学教程》序言中写道："我已经发明了一种进行演示的新方法，简短、容易理解，不需要使用任何语言。"他的意思是，他已经引进了一套完备的数学记法系统。不过，他的完备系统中的符号，到今天还在使用的，只有有关几何的部分：⊥（"垂直于"）和∠（"角"）。

阿尔弗雷德·诺斯·怀特海（Alfred North Whitehead）曾经写道："有一个古老的隽语说，将海洋帝国分配给英国人，陆地帝国分配给法国人，天空帝国分配给德国人。诚然由于来自上天的缘故，德国人接来＋和－；这些符号已经产生的理念，对于人类福祉至为重要，以致不可能来自海洋或陆地。"

字母*p*和*m*取代了文字*plus*（加）和*minus*（减）。普及历史著作将记号＋和－的发明归功于施蒂费尔。但是，也有证据显示施蒂费尔在别处见过那些记号。有人指出，它们最早可能是出现在德国仓库存货单上的粉笔记号，用以表示超过或不足标准重量。

它们出现于1544年施蒂费尔的著作《整数算术》中。它们也出现在威德曼出版于1489年的著作《各种职业中快速且工整的计算》中。但威德曼的＋并非加法运算，而是指"超过"（excess），正如"+2是比预期的还多了2"的意思。其后有一段时间，加法运算的表示法出现了竞争。备受喜爱的有缩写的*p*或$\bar{p}$，直线穿过或置于*p*之上，用以区别是运算符号或表示量的符号。塔尔塔利亚宁愿使用希腊字母ϕ来表示加。代表减的符号要回溯到丢番图的时代，当时这个记号看起来像指上或指下的箭头。水平方向放置的拉丁十字，曾经普遍可见，甚至笛卡儿在他的《几何学》中也偶尔使用铁十字，尽管那可能只是印刷者所加的，他们在自己的字体柜里寻找所能找到的最接近的符号，以避免造新的字模。到了16世纪末，出现了各种代表减

的符号，从÷（我们的除号）到＝（是的，我们的等号）到－，以及最终，实验了其他可能的形式之后，到∿。最后这个∿是埃里冈在他的1634年版《数学教程》中所使用的代表减的符号，这部著作是一本初等数学书，以“几乎不顾一切地引进巨细靡遗的符号”著称。我们今日使用的表示减（法）的水平直线是最简单的一个，但它却导致某种混淆，因为它也在句子中被用作破折号。18世纪之前，这个代表减的符号尚未标准化。17世纪的手稿中，经常在同一页中出现好几种形式的减号。

乘法在奥特雷德于1631年引进符号×之后多年间，并没有固定的符号。哈里奥特使用一个点，笛卡儿则以并列来表示。我们目前仍使用这三种记法。然而，是谁——著作者或印刷商——该负责？我们并不清楚。后来，奥特雷德使用冒号“：”表示除法。分数的阿拉伯符号使用一条线段来分开两个量，其形式在$a-b$，a/b与$\frac{a}{b}$之间变化。我们现在的符号$a\div b$是奥特雷德的冒号与用于分数的阿拉伯线段符号的组合。甚至如莱布尼兹，他日后会创造一些最有意义的数学记法，却使用标记⌣代表乘，标记⌢代表除。我很惊讶它们为什么没有风行起来。它们的设计显得创意十足，因为它们的反射对偶性显示“除”只是“乘”的逆运算。当然，手写是个问题，其中一个会与另一个混淆。

我们现代的无限大符号∞，是古罗马人有时用以表示数字1000的记号，因此是一个很大的数字（参见图4-7和图4-8）。到了16世纪末，这个记号被愚蠢地拿去与雷科德的水平线及克胥兰德的垂直线竞争，看看哪个是代表相等的最佳符号。那个可怜的符号∞被随意摆布，一下子代表这，一下子代表那，一直到1655年，沃利斯在他的《无限算术》中才用它来表示无限大。不过，它还是没有风行，直到1713年，詹姆斯·伯努利（James Bernoulli）在他的《猜度术》（*Ars Conjectandi*）中使用了这个符号。

到那个时候，埃里冈完成六卷著作并在1642年出版，书中的代数以大量的符号来表现。并非每一个人都喜欢这种书写数学的新形式。几部几何著作以几乎不用文字说明的方式印行时，事情看来是走得太过头了。埃里冈宣布他的新方法之后不久，哲学家托马斯·霍布斯在1648年抱怨几何学这种从文字转换到符号的证明方法：

符号是丑陋的，尽管它们是论证的必要支架……尽管它们书写起来比较简短，然而，比起以文字书写，它们并没有让读者更快理解。对于线和图形的概念而言……必须运用口说或思维的文字来进行论证。所以花了双重的脑力劳动，一是简化你的符号成为文字，这种文字也是符号，一是理解它们的意涵。除此之外，你只要考虑没有任何古人在他们出版的几何证明，或者算术书籍中，曾经使用任何符号……我想，将来你就不会如此热衷于它们了。

而德·摩根在1837年写道：

一旦学到符号及其组合法则而不须赋予意义的想法习以为常之后，学生会拥有我将称之为“符号计算”（symbolic calculus）的概念，这种带有某些符号及某些组合法则的东西，就是“符号代数”：一种技术（art）而非科学（science），而且显然是一种无用的技术，除非它将来成为一种科学的语法（grammar）。精通符号计算自然需要意义的赋予。

第18章 趋于标准的符号系统

看似平淡无奇的记法改变，可能暗示着观点上发生激进的转变。任何新记法都可以问出崭新的问题。

——巴瑞·马祖尔

莱布尼兹是一个“中等身材，体形修长，一头棕发，眼睛小却深沉而目光敏锐”的人，一位创造符号的天才。他警觉到使用适当符号的优点，运用符号、改造符号，而当他隐然发现某种设计不佳的符号有一天可能让数学解说不必要地变得复杂时，则会抛弃那些符号。他研读了邦贝利和韦达的著作，并预见到多项式的符号在17世纪初代数一般化后将不敷使用。他了解，使用起来不方便的符号如何在15世纪和16世纪让代数的发展陷入困境。

到了17世纪后半叶，数学手稿充斥着符号，这主要归功于莱布尼兹，以及奥特雷德、埃里冈、笛卡儿和牛顿。教科书作者和一些鲜为人知的数学家，设计出数以百计的新符号。在那个时代，创造符号风行一时，但对于预料之外的混乱缺乏了解，迟早会使富有创造性的想法陷于绝境。

如同我们所见，笛卡儿借用了极多符号，并调整改进它们。奥特雷德未多加考虑一些符号的价值，即引进了数百个可能有用的符号。即便有些符号明显有问题，奥特雷德仍为了保持稳定的一致性而继续使用它们。埃里冈也是如此。

而莱布尼兹想让他书写的内容更清晰明确时，会优先使用符号。他相信卓越的记法是理解所有人类思维本质的关键。“真正的方法，”他写道，“应该有一条阿里阿德涅的线（Ariadne’s thread）带领我们前进，也就是说，有一种切合实际又极其明了的工具，它会引导思维，就像几何学中所画的线和为学习算术者制定的运算公式的功用。”

在希腊神话中，阿里阿德涅是克里特国王米诺斯的漂亮女儿，忒修斯是从雅典送来献祭给迷宫中的米诺陶（半人半牛的牛头怪）的青年。阿里阿德涅爱上了忒修斯，给了他一个线球（clew），让他在进

入洞穴时把线球松开。一旦杀死了米诺陶，就可以循着这条线走出洞穴。忒修斯成功杀死米诺陶并离开洞穴后，带着阿里阿德涅前往纳克索斯岛，并将她抛弃在那个岛上。

clew这个词原意是“线球”（至今仍是），后来演变成我们现在所用的词clue（线索）。显而易见，莱布尼兹用“阿里阿德涅的线”这个象征来传达一种想法，也就是数学和其正确推理的力量的线索，存在于其记法的特性当中。莱布尼兹的微积分记法如此完美地符合这门学科的基本逻辑运算和程序，一般学生能够循着这条线穿过推理的迷宫并找到出口，确信自己理解了内容而受到激励。

莱布尼兹了解符号，它们在概念上的力量及它们的限制。他会耗时数年进行实验——创造、否决、调整了许多符号，并用符号来与自己认识的每一个人交流，还请教了当时多位一流的数学家，那些数学家对莱布尼兹的一丝不苟非常有共鸣。莱布尼兹没有轻易使用雷科德的等号，而大半偏好用一个看起来像订书钉的符号“$\sqcap$”来代表等于。我想是因为这个符号暗示它是连接两边的桥。

依照惯例，我们现在说“y是x的函数”，并用$y=f（x）$的记法来表示：对于每一个x值，f都能将之对应到一个特定y值。1692年，莱布尼兹撰述关于曲线切线的内容时，提到了一种更严格限定的观念。对他来说，函数只是代数与分析学的运算中所造出的一种表达式——例如，$ax+b\sqrt{a^2-x^2}$具有函数的资格，因为它是由加、乘、指数和开根号等代数运算建立起来的。1837年之前，函数的概念历经多次修正，直到狄利克雷为函数所下的卓越定义在那一年确定下来，而那也是我们今日在数学中使用的函数定义：“若每一个x值，都有唯一一个与之对应的y值，则y是x的函数。”狄利克雷的定义放开了函数在求对应值时的所有限制。笛卡儿不用前述那样自由的定义；他必须将方程式对应至曲线，才能比较容易地探究一个变量如何随着另一个变量移动，就像空间中的点随着时间移动一样。

在莱布尼兹发明的两百多个新符号中，包括他的微积分的积分与微分的符号。学过微积分的人都见过符号$\dfrac{dy}{dx}$，也就是“y对x的导

数”（参见附录A）。

为什么$\frac{dy}{dx}$是如此优良的符号？若不质疑符号操作的不合理性，$\frac{dy}{dx}$可能被看成是一个分数；我们可以在一个方程式比如$\frac{dy}{dx}=x$的两边乘上dx，得到$dy=xdx$。多么方便啊！结果证明，那些奇怪的小变量dx和dy，综合来看的确遵循了代数法则。

莱布尼兹的微分符号dx，dy和积分符号$\int$，比其他研究微积分的数学家所用的任何符号都好得多。相较于使用牛顿或费马留下来的符号，莱布尼兹的符号让有关微积分的工作变得容易得多。排字工人反对像$\frac{dy}{dx}$这样三层的符号，这种符号会打乱页面上的行距。我们差点被莱布尼兹的另一个符号$^{1}\!dy$缠上，这个符号看起来像d的上部断开往左移，或者把那个断开的部分想成上标1。我想这样可笑的符号会成为排版者的梦魇。幸好这个符号已不再使用。

这样的排版考虑对符号设计影响很大。莱布尼兹遵循使用群组线的惯例，把会被共同运算的项集中在一起；群组线延长盖住那些被集中一起操作的项。但它也造成了排字工人的困扰，所以莱布尼兹发明了另一种不会让页面行距变宽的方法。于是，为了满足排版者，并且让版面看起来更赏心悦目，他借鉴了使用一对括号的想法，来表示哪些项要被集中在一起。

莱布尼兹非常确信符号改良很成功，夸耀道：

> 我认为，当这项工作完成后，它会是人类心智的最终成果，而众人皆将欢喜，因为他们将拥有一项得以颂扬智识的工具，如同望远镜让洞察更臻完美。

第19章 站在巨人肩膀上的侏儒

牛顿是一个“外表和举止都非常没有活力，那些不认识他的人不会对他有多大指望”的人，对于那些他站立在其肩膀上的巨人，他给了比喻性的赞美。牛顿广为人知的名言：“如果说我看得比别人更远，那是因为我站在巨人的肩膀上。”这句话可回溯至12世纪新柏拉图学派的法国哲学家伯纳德（Bernard of Chartres），他将自己那个时代比喻为“站在巨人肩膀上的侏儒”。伯纳德指出，我们比前人看得更多更远，不是因为我们有更敏锐的洞察力或站得更高，而是因为“我们被抬至高处，站在巨人高大的身量之上”。

20世纪的数学家及研究牛顿的史学家赫尔伯特·特恩布尔（Herbert Turnbull）说了一段少年牛顿的故事，有趣又富想象力：

> 大约在克伦威尔去世之时，格兰瑟姆（Grantham）附近的乡村正经历狂风暴雨，一个男孩以奇妙的方式寻找乐趣。他转身背着风，跳了一下，当然他跳得很远。接下来他又转身面向风，再次跳了一下，这一次不像第一次那么远。他仔细测量这些距离，用这样的方式来弄清楚风力大小。这个男孩是牛顿，而他后来测量了一种力，一种让行星在其轨道上运行的力，如果这种力真的是这么回事的话。

到了那个时代，望远镜已臻完备，探险足迹遍及世界各大洋，然而，女巫仍然会被吊死或烧死，叛国贼和罪犯在公共广场上被斩首是家常便饭——他们的头经半熟防腐处理后挂在沿街的柱子上。另外，即使最聪颖的科学家，在面对崭新的化学科学时，也成了炼金术的狂热拥趸。连牛顿都坚定地投入炼金术实验。

17世纪和18世纪的数学家从坚持数学的严密性的古典希腊传统中适度解放，在他们关于无穷大和无穷小的直觉和猜测中获得力量。有了无限，数学家必须发展出新的法则和新的记法。过去定义模糊，方法不明确，而逻辑论证对不完整的逻辑链妥协。丹齐克写道：“直觉已经被希腊人苛刻的严密性禁锢太久了。现在它无拘无束，而且没有欧几里得们抑制它的幻想驰骋。”

无穷大和无穷小量的工具，连同对连续统的直观理解，都被创造出来并获得接受。虚数成为数学名词。代数和其符号的巧妙使用，让

数学准备好迎接微积分革命。物理被提升成为一门科学。而牛顿已经——引用爱因斯坦的话——“完成了任何个人能够成就的最伟大的思想跃进”。

多亏了英国数学史家德里克·托马斯·韦赛德和他的辛苦编辑，我们拥有几乎牛顿所有的论文。1958年，汤姆（朋友对韦赛德的称呼）首次研读牛顿的论文，当时这些论文一团乱。身为剑桥的研究生，他的论文是研究17世纪的数学史，他开始意识到现存的多数数学史可疑又没有条理。据说他问了剑桥一位图书馆管理员，有没有牛顿的手稿可借阅，然后很快借到了八箱。汤姆花了二十三年完成八大册牛顿手稿的编辑工作。

我不时会借阅第七册，似乎有只老鼠与我共享那本书，它在装帧上发现了某种不知道是啥的美味东西。我一次只读一页，光是这第七册就满载足够我用尽一生去思索的信息。

牛顿为沿着曲线流动的量构想出未知的变量。他称它们为*fluent*，源自拉丁文*fluxus*（“流动的”），非常接近我们现在所称的**因变量**，也就是x，但它们受限于对时间的依赖。

1704年，牛顿首次使用这些变量近四十年后，对它们的描述如下：

> 我这里所考虑的数学量，并非由最小的可能部分组成，而是由一连续的运动来描述。线的描述，并非由部分的并置（apposition）所产生，而是由连续运动的点来描述；表面积是由线的运动来描述，体积是由表面积的运动来描述，角是由边的旋转来描述，时间是连续的通量，依此类推。这些皆起源发生于真实的物理本质中，并日复一日在物体的运动中得到印证。

这个观念与莱布尼兹的数学量多么不同呀。对莱布尼兹来说，曲线是固定的、静态的，以其方程式来描述，由具有无穷小的边的一个无限多边形组成。另一方面，牛顿认为曲线是动态的，是一个移动粒子的轨迹，在这条曲线上，任意切线都会指向粒子飞出曲线的路径之瞬间方向。他认为曲线就像是“点的流动”，代表了量；但在微积分里，它们相当于莱布尼兹的静态曲线。

随着时间的改变，曲线上的量沿着曲线衍生出一个新的量。流量改变的速度就是“流量的流数”，一个冗长拗口的词，写成单点在上

的符号$\dot{x}$，$\dot{y}$，$\dot{z}$——所谓标点字母，数学界很快就接受这些符号成为标准的**流数**记法（参见附录B）。令人好奇的是，高次的导数会在变量上方记上多个点，所以$\overset{\cdot\cdot\cdot\cdot}{\underset{\cdot\cdot\cdot\cdot}{y}}$就代表$y$这个流量的八次流数，意思是将流量$y$取流数之后，再取流数……（共八次）；在任何人提出写成像$\overset{8}{y}$这样的构想之前，似乎得把以前指数指标的故事再说上一遍。我们现代的莱布尼兹记法将此记为d^8y，这种表示法让人满意得多。只要想象一下得写$ddddddddy$来表示微分八次，就可以了解了。在莱布尼兹的时代，对这么高次的微分可能没有迫切的需求，但更复杂的项无可避免地终将出现：$12ddddddddxdddddydddddz$这种梦魇仍旧会现身。幸运的是，莱布尼兹的记法写成：$12d^8x \cdot d^5y \cdot d^5z$。

下一个问题是，由于记法的限制，对于流数的理解，需要一种语境来让自变量在概念上的本质意义更清晰明了，这种自变量通常是时间变量t，但非必然如此。x的流数被理解为与时间变量相关，所以不过是代表x的速度罢了，用莱布尼兹的记法，这写成$\frac{dx}{dt}$。用牛顿的记法，则记为$\dot{x}$。而以现代的语言来说是：x对t的导数。

根据牛顿的观点，微积分的基本任务，还是要用一堆“流”来解释：对于给定的流量，可找到对应的流数，以及当给定流数时，可反过来找到对应的流量。然而，牛顿一生中运用了好几种方法，同时也支持无穷小量的研究。

以$y-x^2=0$为例，我们可以用$x+\dot{x}o$来替代x，以$y+\dot{y}o$替代y（牛顿求x^n的流数的步骤参见附录B）。o是字母“o”，指一个非常非常小的量，但不是零。事实上，o代表牛顿所称的无穷小量，不管它原来指的是什么。在此条件下，前述方程式变成

$$y+\dot{y}o-(x+\dot{x}o)^2=0$$

并等价于

$$y+\dot{y}o-x^2-2x\dot{x}o-\dot{x}^2o^2=0$$

因为 $y-x^2=0$，上式变成 $\dot{y}o-2x\dot{x}o-\dot{x}^2o^2=0$。[1]牛顿认为$o$很小，但不是0，因此除以$o$完全可行。除以$o$后，上一个式子变成 $\dot{y}-2x\dot{x}-\dot{x}^2o=0$。然后牛顿会论证一个值得商榷的问题：因为$o$表示一个无穷小量，乘上$o$的项与那些没有乘上$o$的项相比，差异微不足道。因此，牛顿拿掉$\dot{x}^2o$这一项，所以上述最后一个式子变成$\dot{y}-2x\dot{x}=0$。当$o$不是零时，除以$o$没有问题，完全可行。然而，当讨论到$o$不是零而是无穷小时（不管无穷小是什么意思），就为哲学家及爱尔兰科克郡克洛因（Cloyne）圣公会主教乔治·柏克莱（George Berkeley）留下了一些悬而未决的重要问题：

> 的确，必须承认，他使用流数，就像建筑物的脚手架一样，一旦有限的线段被认为与它们成比例时，流数就被弃置一旁或丢掉。但接着，这些有限指数借流数之助被求得。……而这些同样转瞬即逝的增量究竟是什么呢？它们既不是有限量，也不是无穷小的量，但又不是零。我们能不能称它们为消失量的幽灵（the Ghosts of departed Quantities）？

牛顿希望这个称为无穷小的东西有双重用途，而莱布尼兹对这个问题的看法与牛顿相同：它不是0，可以拿来当除数，但某种程度上又像0一样可以忽略——“消失量的幽灵”。

根据柏克莱的看法，牛顿的微积分未能符合连续性的直观概念。仅从柏克莱的论文副标题便可看出他的观点：或一篇致一位不信神数学家的论文。其中审查一下现代分析学［指微积分］的对象、原则及论断是不是比宗教的神秘、信仰的要点有更清晰的表达，或更明显的推理。

真正的争论在于，当分母和分子都趋近于零时，牛顿有关比的极限的含混意义的证明问题，那是一种忽略微妙差异之处及无限和连续性的棘手问题的精巧观念。牛顿没有把那些比当成真正的比，而是作为极限，如同我们今日的看法。对柏克莱来说，牛顿似乎用零去除零，是毫无意义的愚蠢行为。

这位主教的抱怨是可以理解的，对于欧拉、费马、牛顿和莱布尼兹这些有着良好直觉的数学家来说，直觉是好的。危险之处在于，一些狡诈无序的想法可能伪装成一项已经定明的定理的合理产物，偷偷溜进微积分的大门。到了18世纪末，微积分和坐标几何的实际应用与日俱增，改善了现实世界的人类生活和知识发展，不理会那些从推理大门偷溜进来的矛盾事物。微积分的发明促进了建筑、天文学、火炮技术、木匠技术、地图学、天体力学、化学、土木工程、时钟设计、流体动力学、流体静力学、磁力学、材料科学、音乐、航海学、光学、气体力学、造船和热力学的发展——而这绝对称不上是详尽的清单。

到了1727年牛顿去世时，单片眼镜和报纸很容易买到且价格负担得起。政治剧变笼罩欧洲；欧洲中部的小公国于战争中合并，成为新王国，同时波兰和奥斯曼帝国被邻国瓜分了大片领土。城市人口仍然不多——伦敦不到六十万，巴黎不到七十万——还有狼群在城市外面随意漫游。明亮的咖啡馆和奢华的周遭环境，在欧洲各大城市和大学城里随处可见，每天下午都有报纸销售，晚上街道明亮，人们可以散步及讨论政治、哲学和最新的科学发现。当时欧洲正在体验一种崭新的生活风格。咖啡馆不仅是闲聊和传递消息的地方，也是学生和教职员谈论书籍、诗歌和剧作，收信或听听最新科学传闻的场所。科学院和科学学会创建起来，有资金出版定期刊物，也有经费发展研究工具和昂贵的测量仪器。

牛顿去世后五十年间，狄德罗完成了共十七大卷的第一部百科全书，吉朋以《罗马帝国衰亡史》震惊了世界，卢梭写出《社会契约

论》，瓦特建造了蒸汽机，莫扎特谱出小夜曲和交响曲，巴哈去世，而贝多芬诞生。

虽然奴隶交易增加，战争也从整个欧洲的国家蔓延到殖民地、贸易和海上霸权，但科学、艺术、文学和各种实用的发明，即将在启蒙时代迅速发展。中产阶级变得见闻广博并开始思考，不仅局限于政治，也包括科学和文学。

全球化信息高速公路已经准备好，可以传播灾难、流行知识和科学发现的消息。人类文化的变动不断急剧复杂化，而且即将带来伟大的发现，但行星的运动，更别说炮弹和箭矢的轨迹，似乎基本上都需要用微积分来确立。这是一个见证科学全力冲破界限的时代，当时教科书作者正寻求新方法来向日益增多的大学生说明数学。

---

[1] 原文有误，兹订正之。——译者注

第三部分 符号隐藏的力量

好奇的读者可能想知道，

被大量使用的符号背后深藏了什么秘密，

以及符号成为今日我们所见形式的演变过程。

那些特别的瞬间对现在的我们来说似乎显而易见，

但对过去的思想家而言却是巨大的飞跃。

第20章 只可意会不可言传

一个好的记法可以把大脑从所有不必要的工作中解放出来，让大脑专注于更高深的问题……

——阿尔弗雷德·诺斯·怀特海

16世纪之前，几乎任何人只要有足够的决心，都能理解几乎任何数学著作的重要内容。准备好鹅毛笔和羊皮纸，一个安静的房间，打开窗户让清新的微风吹拂，足够的兽脂让蜡烛燃烧竟夜，再毫无顾忌地让思维纵横驰骋，就可能以自然语言写出数学著作。对任何希望解析数学的语法、动力、机制和逻辑的人来说，数学都是可阅读的。

《爱丽丝漫游奇境》中的《无聊诗》（Jabberwocky）开头是“备餐时光，软黏托佛”（Twas bryllyg， and the slythy toves）[1]，给人留下一种印象，感觉像是对门外汉来说听起来艰深但合理的语言，就好像婴儿刚开始理解周围的声音时听到的感觉。听听接下来的诗句“在伟伯上仪转及锥钻／贫弱到底的波若哥夫斯；／迷途瑞斯的吼哮”（Did gyre and gymble in ye wabe / All mimsy were ye borogoves; / And ye mome raths outgrabe），隐然觉得似乎有些道理。当我们第一次遇到自己不了解的数学，或任何事物时，我们会像在读《无聊诗》。

到了18世纪，数学语言对于没有受过大量预备指导的人来说，读起来已经太过符号化。这不是因为符号的数量增加了多少。数量不是问题，而是初学者尝试理解新材料时，必须学习一种新的视觉语言。要理解这样的语言，需要运用非常特殊的专业知识或坚持不懈地工作。这个语言是看得见的，但其意义隐藏其中。包装在记法句子中的符号提供了大量信息，只有那些有时间、有天分或有耐心接纳这些信息的人，才能了解它们的内容。

当人们不懂某件事时，我们经常听到他们调侃说“简直像希腊文一样”（It’s Greek to me）。希腊文不是特别难的语言，希腊宝宝学习希腊文时，就像美国宝宝学英文一样容易。那么，为什么专挑希腊文来表示不理解呢？很有可能是因为希腊文写起来不像西方人非常

熟悉的那些拉丁字母。那种对希腊字母的陌生感，本身就带着一种自我膨胀的无知。

数学符号原本是用来帮助我们理解的。它们意在帮助我们了解数学论证，让事情变得简单，慷慨地赋予我们精简的形式，以便我们在阅读数学时，对于究竟是怎么回事多少拥有一点直观印象。但就像我们难以理解的任何术语一样，数学真的像令人泄气的希腊文——部分原因是它们往往**真的是**希腊文。

怀特海挑战我们说：

> 如果有任何人怀疑符号的实用性，请他在不使用任何不管什么符号的情况下，一字不漏地写出下列等式的完整意义，它们表示的是一些代数基本法则：
>
> $$x+y=y+x$$
>
> $$(x+y)+z=x+(y+z)$$
>
> $$x\times y=y\times x$$
>
> $$(x\times y)\times z=x\times(y\times z)$$
>
> $$x\times(y+z)=(x\times y)+(x\times z)$$
>
> ……这个例子说明，借由符号体系之助，我们几乎可以凭眼见即机械性地进行推理的转换，否则就需要发挥大脑更高的功能。

尽管托马斯·霍布斯称符号为“论证的必要支架”，但他也写道：“相较于用文字书写，它们完全无法让读者更快了解［被符号化的东西］。”

> 对于线和图形的概念而言……必须运用口说或思维的文字来进行论证。所以花了双重的脑力劳动，一是简化你的符号成为文字，这种文字也是符号，一是理解它们的意涵。

在自然语言里，即使最精心选择的文字也会带着隐而未显的意义，那些意义具有操控推理的力量。我们从词典里学到一些词，而它们通过我们已经知道或我们可以查到的词来给出意义。我们理解（并学习）其他词的方式主要是通过适应含糊的意义，判断那个词的各种可能意义在不同语境下如何使用才是最好的方式。椅子和凳子、杯子和马克杯、门口和玄关，这些词之间有何不同？

数学符号有时也像这样带有隐而未显的意义，但它们的目的是带来纯粹的思维。从语境中了解一个数学符号代表什么是可能的。我们往往是从数学符号的定义来了解那些符号的意义——我们说**往往**，因为在正规数学中，不是每个人都能轻易领悟那些与熟悉的经验特质无关的定义。华威大学的大卫·拓尔（David Tall）和希伯来大学的什洛莫·文纳（Schlomo Vinner）在一篇划时代的论文中指出，许多待定义的概念，在给出任何正式定义之前，已经以个人图像的某种认知结构存在于头脑中。

符号语言当然促进了它本身隐而未显的意义，这些意义源于进入潜意识的富有想象力的灵光闪现，但那些最好的符号可以精准指明意义又让头脑得以在它类似语境模式的数据库中，快速来回比较、传递，以富有创造力的方式将未知的与已知的联系在一起。

数学使用符号是为了精确表达内容。达西·汤普森在其经典名著《生长与形态》（*On Growth and Form*）中问道，我们如何分辨彩虹的形状与软管喷出的水柱弧线的差别。它们看起来可能一样，甚至可能都出现七色彩虹。两者都是由水滴构成的。以日常语言来说，你可能会说它们是看起来很像的平滑弧线。

但通过符号的透镜来细看那些曲线，它们的形状大不相同。彩虹上每个坐标点（x，y）都满足方程式$y=\sqrt{a^2-x^2}$，而水流弧线上的每个坐标点（x，y）则满足方程式$y=ax^2+bx+c$，其中a，b和c是决定曲线起点与终点间的高度和宽度的固定值。一个是半圆，另一个是抛物线。不管如何修改参数a，b和c，这两条曲线都不可能重合变成一条曲线。

有了便利且适当的符号，我们就能专注于可能显现出来的模式、对称、类似和差异，这些在自然语言的透镜下会模糊不明。

以方程式$x^2+y^2=xy+4$为例。嗯……如果xy项不在就好了，这样我们就会得到一个半径为2的圆，也就是$x^2+y^2=4$这个方程式描绘出的圆。但xy项**在**那里，那它是如何改变那个圆的呢？如果不以某种变换来简化这个方程式，无法分开两个以此方式纠缠在一起的变量x和y。然而，原始方程式中x和y的对称性，为我们提供了曲线几何的线索。如果将x和y交换，你会得到一模一样的方程式。啊哈！这不就表

示说，这是对称于$y=x$这条直线的曲线？的确，将坐标轴顺时针转45度，用s和t来表示新的坐标轴，原来的方程式神奇地变成了$3s^2+t^2=8$。这个新的形式没有st项，亦即s和t没有纠缠地相乘在一起。在s和t坐标上画出这个漂亮方程式的图形，那是一个以（0，0）为中心、对称于s轴和t轴的椭圆。

就像方程式$x^2+y^2=r^2$的对称形式大喊着：圆！圆！xy这个以乘法将x与y结合在一起的项也是，它马上对着我们的左侧大脑皮质尖叫：旋转！旋转！通过让xy项消失，旋转45度解决了变量x和y的问题。

方程式中的对称，往往意味着可以通过那个方程式描述出某种在几何上对称的曲线。而这当然适用于我们的方程式$x^2+y^2=xy+4$。这个曲线是椭圆，对称于两条与水平轴呈45度角的斜线。

头脑可能无法很快判定这条曲线是个椭圆，但这个方程式迅速告诉我们：不管它是什么，它必须对称于$y=x$的直线，因为x与y交换不会改变那条曲线，所有会改变的只有变量的名称。

把代数与几何联系在一起的潜在力量牢不可破，却几乎难以察觉。这些潜在力量使得代数的程序清晰可见。它们在我们的思维中赋予我们模式、联想、类似和不可思议的交会，而这些如果仅用文字会让人如堕五里雾中。接受怀特海的挑战，试着不使用任何不管什么符号，一字不漏地写出$x^2+y^2=xy+4$这个等式完整的几何意义。这当然可以做到，但是你的血压会爆表。

对称有许多不同的形式。一个老掉牙的好笑问题问“乔治·华盛顿的白马的颜色”，它确实诱使我们去探查问题本身再去找答案。当我们问一个平方是4的数的平方为何，我们就提问了一个反身自答的问题。用符号来表示，这个问题即是答案，答案即问题：$(\sqrt{4})^2=4$。表面上看，这种同义反复的等式并未问出新信息，也未问出任何信息。但如果我们用符号来表示，将它一般化到所有的正数，就像$(\sqrt{x})^2=x$，我们富有创造力的天赋

激发出类似的问题：$(\sqrt[3]{x})^3 = x$这个等式为真吗？$(\sqrt[4]{x})^4 = x$呢？对任意正整数n，$(\sqrt[n]{x})^n = x$呢？

由此，我们的天赋可能跃进到一种对符号$\sqrt[n]{x}$的全新理解。若x^n代表x自乘n次，且若对正整数n和m，$(x^n)^m = x^{n\times m}$，则用符号$x^{\frac{1}{n}}$来表示$\sqrt[n]{x}$不是很合理吗？“该数的n次方是x”，假设这样的东西真的是个数的话。这样一来，代数证实了我们已知的东西，但也让它本身扩展到包含一种指数的算术。我们会得到：

$$(x^{\frac{1}{n}})^n = x^{\frac{1}{n}\times n} = x^1 = x$$

从上述等式，小小跃进到了解当n和m是正整数时，$x^{\frac{m}{n}}$就代表$(\sqrt[n]{x})^m$，定义也小小飞跃到将n和m扩展至所有的整数。而由此，利用定义、推理和模式来导引出更有力的通则，了解符号体系和定义如何从一个聪明的想法到逐一建立下一个想法。

取幂和开方（taking powers and extracting roots）这美妙的符号形式，提供给我们指数算术的语法。因为它们的定义方式，取幂和开方互为逆运算。加和减互为逆运算，加一个数再减同一个数，会得到原来的数。乘和除也是同样的道理。一般来说，一个数学运算有逆运算时，它会很有用。这样的逆运算是解方程式的关键。举例来说，要解方程式$x+2=4$，我们在两边各减2，得到$x=2$。要解方程式$x^2=4$，在两边各取平方根，得到$x=\pm 2$。

数字已经进展到远远不同于计数那些早期的开端，计算出我们有十根手指、两只眼睛和一个鼻子。它们不再专指我们看到的东西，或

我们需要计数的东西。现代的数学感兴趣的是看起来用符号来表示的定义、推理和模式。只要遵循了数学法则和符号语法，定义甚至可以与日常用语及直观概念抵触，特别是直观的物理概念。它们打开了通往可见的自然之外的逻辑世界的大门。没有什么比你开始想有理数之外的东西是什么更明显的进展了。唯有纯粹的数学语言，带着它高度发展的符号意识，能看见可见的东西之外是什么。

曾经，可能存在一个数而它的平方是负的这个观念，似乎远远超乎人们的想象。$\sqrt{-1}$这种想象的东西到底有啥用？用正确的符号语法来解方程式$x^2-2x-2=0$，你会求得两个合理的解：$1+\sqrt{3}$和$1-\sqrt{3}$。但如果你试着用同样的符号语法来解二次方程式$x^2-2x+2=0$，会出现什么事呢？出现两个奇怪的解：$1+\sqrt{-1}$和$1-\sqrt{-1}$。取其中任一个解，将其平方，减去本身的两倍，再加2，得到0。这些解个别看起来可能是无效的，但把它们相加，得到的只会是2。换言之，这个奇怪的项$\sqrt{-1}$在代入方程式的过程中消失了。

如果你刚刚来自16世纪早期，正在阅读本章，你可能会想到——你也应该想到——将$1+\sqrt{-1}$和$1-\sqrt{-1}$相加得到2，是一件可疑的事。这样的和意味着$\sqrt{-1}-\sqrt{-1}=0$。这为真吗？现代的答案是：“当然为真，$x-x$必须等于0，不管x是什么。”这件事对遵循一般算术法则的数当然是正确的。但到目前为止，我们只知道$\sqrt{-1}$是源自将符号代数用于求解二次方程式之后所得的符号。我们知道$\sqrt{-1}$代表神秘的**某物**，而这个某物确定具有下列性质：“如果你把它自

乘”，无论那是什么意思，“你会得到负数−1”。除此之外，我们对它一无所知。

你——来自古代的陌生人——可能会认为$1+\sqrt{-1}$和$1-\sqrt{-1}$是没有意义的解，因为没有任何明显的事实告诉你，真实世界有什么现象可以导出像$x^2-2x+2=0$这样的二次方程式（以今日的记法表示）。如果你来自16世纪末，并且了解一些关于二次方程式在直角坐标系上的图示，你可能会说那个图形是一个抛物线，其最低点是（1，1），在x轴上方两个单位的点。没有x使得y的值为0。

但思考得远一点。摒除直角坐标系，想想不同的情况。如果我们现在将所有形式为$a+b\sqrt{-1}$的数纳入我们的数系，其中a和b可以是已经纳入实数俱乐部里的任意数，会怎样呢？你或许认为纳入那些数没有意义；然而，以符号的方式来看，不管这些东西是什么，它们在我们普通的数的语法（grammar and syntax）内的运作没有瑕疵。它们似乎遵循普通数的所有法则：两两相加、两两相减、两两相乘、两两相除，会得到形式为$a+b\sqrt{-1}$的另一个数。相信所有的惯用法则都适用，然后做运算吧！但是为什么不像1，2，3……这类正规数，也不像稍微有点奇特的数$\frac{3}{4}$，π或$\sqrt{2}$那样，除了你开始习惯符号$\sqrt{-1}$的意象之外，你心中没有任何这个数所可以代表的清晰意象？

你可能认为$\sqrt{-1}$是早已存在的一个代表−1的平方根的符号，但它不是为了这个目的而造出来的。它是从尝试求解方程式的代数运算的结果中产生的。它看起来好像是运算过程中偶然将根号套在了负数−1上。不过，运用一点算术的巧妙手法避开细节，注意到形式为

$a+\sqrt{-c}$的任意数，其中a和c为实数，都可以写成$a+b\sqrt{-1}$，其中a和b为实数。因此，$\sqrt{-1}$展现了实质的重要性，所以它值得拥有一个特殊的符号。我们用字母i来表示它，灵感来自imaginary（想象的、虚幻的）这个词。以bi形式表示的称为“虚数”，其中b为实数；以$a+bi$形式表示的称为“复数”，其中a和b为实数，“复”这个字的意思是表示混合实与虚。（不幸的是，“虚”和“复”这两个字在数学词汇里根深蒂固。之所以说**不幸**，是因为这两群数既不虚幻也不复杂。）

但令人意外的或许是，符号i（虽然它只是imaginary这个词的缩写）明显优于$\sqrt{-1}$。阅读数学时，$a+b\sqrt{-1}$之间的差异，就像捏着鼻子吃草莓闻不到香甜的滋味，与正常呼吸吃草莓的差别。

数？为什么我们称这些东西为数呢？我们曾经认为数应该是某些东西的**计数**——手指、脚趾、羊、日子、德拉克马（drachma，古希腊银币）、眼睛、耳朵、鼻子。然后，我们认为数是某种事物的**度量**，可能是分数或无理数。但这些所谓的复数怎么用来计数或度量？也许我们应该称它们为“成对的数”，但就算这样，还是无法满足我们一般对何谓数的判断。它们甚至不是成对的数，因为有个麻烦的东西附加在第二个数上。

我们可以将整数想象为一条线。从0开始量单位距离，以这种方式把每一个整数定位好，正的整数放在右边，负的整数放在左边。这也适用于有理数和实数，它们可以写成无穷小数。在累积的文明意识中，我们似乎不断在追求数在头脑中的图像，即使那个图像是模糊不清的。但要让复数可视化，需要某种更富创造力的东西（参见附录D）。

数的概念曾经被表示为一个简单的形容词：“十（根）”手指。过了很久很久，它变成一个名词“十”，无关乎具体的量词。但自16世纪中叶后，符号大量涌入数学的语言当中，在概念上，数的定义被

扩大，以便含括一种动作（act）或一个存在（being）的模式。现在我们有了i，代表一种动作的一个数——旋转90度的动作。

关于那些复数：回想卡丹诺用来解三次多项式的公式带来虚数可能有作用的想法——尽管它们的名称让人觉得有点可惜——那时即使最权威的数学家和哲学家也被它们的神秘难解迷惑。部分原因归咎于它们显而易见很难应用。玫瑰如果改成任何其他名字，它将依然芬芳，那么$\sqrt{-1}$不叫作“虚”数而改称其他名称，也将依然真实。这是个倒霉的名字，不管它虚不虚。

负数是处理$x+a=b$这类方程式时出现的，虚数则来自$x^2+a=b$这类方程式。所以，一个方程式是否有道理仰赖于符号a与b的关系，以及负数平方根的正当性。当a大于b的时候，产生一个问题：有一个数自乘之后，结果是负的。这根本没道理，除非我们将负数的平方根纳入数字俱乐部。为了让这个没道理的东西变得有道理，我们必须回到数学的根本问题——何谓数？——使得所有形式为$x^2+a=b$的方程式的解都有意义。我们希望$\sqrt{b-a}$有意义——**永远**——只要a和b为有理数。

也许，这个荒唐的符号i，会以某种神秘的方式，拥有某种正当性，某种比虚幻来得真实的东西。也许那些奇怪、无意义的符号可以通过某种方式，导出问题的解答并产生有效的结果。

怀特海曾调侃道：

> 一个符号未经恰当地定义，根本不是一个符号。它只是纸上的一滴墨渍，有个容易认出的形状罢了。一串墨渍证明不了什么，只能证明有支坏掉的笔或一个粗心的作家存在。

复数$x+iy$最终变成了只是必须遵循一堆简单法则的实数对（x，y），它在解决流体流动、热传导、重力和几乎整个数学物理的问题上，都作用惊人。复数对的加和乘法则的图形表示出奇简单，而这样的运算的意义同样出奇简单。

一件关于数学的奇妙事情是——借由它的最佳符号——它的进展开阔了它的视野。任意实数乘上−1，就可以让每一个正数变负，每一个负数变正。从实数线的图示来看，你已经将原本的实数线整个旋转了180度，原本向右递增的数变成向左递增。在二维平面上，任意复数乘上i，就会让它逆时针转90度。

当你试着根据三数组（x，y，z）建一个三维数系时，无可避免地最终得到的数系会有个称为“零因子”（zero divisor）（非零的数相乘为零）的讨厌东西，这让用来解方程式的正规代数一团混乱。因此，略过三维空间，直接到四维，这下一个维度可以形成一个遵循结合律的数系——其中a·（b·c）=（a·b）·c，没有零因子。当然，这需要付出代价：我们必须要放弃交换律——a·b不再等于b·a，不同于截至目前我们所碰到的所有的数的情况。

这种“四元数”（quaternion），如19世纪爱尔兰数学家哈密尔顿所称呼的，属于四维中的一种新数系，包含复数和一个乘法系统，遵循除了交换律之外的所有代数法则。哈密尔顿和妻子在都柏林散步时发现了这种四元数。“当下，我感觉思绪的电路突然中断，”他写道，“而从那里迸出的火花就是i，j，k的基本方程式，此后我一直应用它们。”（更多关于四元数的讨论，参见附录E。）重新斟酌怀特海提出的挑战：试着不使用任何不管什么符号，写出四元数基本方程式的完整意义。

---

[1] 此处中译文引自林望阳译《镜中奇缘》，台北小知堂文化，2000年。——编者注

第21章 符号背后的意义

1706年，符号 π 首度现身。威廉·琼斯（我们多少人曾听说过他？）用希腊字母 π 来表示圆周长与圆直径的比，多简洁呀。“对于将这位来自希腊字母领域的尊贵访客带到数学史舞台上，没有为读者准备冗长的介绍词。它只是来了，无人宣达。”但接下来三十年，没有再使用这个符号，直到欧拉在与詹姆斯·斯特林（James Stirling）的通信中用到它。

我们可以指控 π 不是一个真正的符号，毕竟它只是periphery（圆周）这个词的第一个字母。没错，但就像 i 一样，它使人想起一些概念，而这些概念在使用有太多包袱的符号时可能不会显现。某些像“什么是 i^i”这样的问题，可能未多加思索就脱口而出。纯数学探究这样的问题，是因为它不仅致力于符号化的定义和法则，也通过提出日常用语可能忽视的问题来看看数学的边界可以推得多远。你可能认为 i^i 毫无意义，认为那啥都不是，或可能是个复数。令人意外的是：原来它是个实数！

相较于我们第一次开始数草地上的羊时，数似乎有了一种更广泛得多的意义。我们扩展了想法，纳入各种各样的概念性事物，包括仍遵循数值运算法则的数字家族的那些不寻常的数。就像我们使用的许多词语一样，数所具有的意义已经比它曾经代表的意义更广泛得多。

马赫深思说：

> 只要想想那些数学家长期进行运算的所谓虚量，而且，在他们有立场对它们赋予一个完全确定且又能够可视化的意义之前，他们甚至已经从它们取得重要的成果。

坚持真实世界的相关性并不是数学家的工作，但这个世界似乎最终还是采用了数学的抽象概念和普遍原理，并将它们应用到某些与这个世界的存在相关的事物上。几乎整整一个世纪，数学家使用虚数的指数之后，一个新的概念萌芽了。于是，从曾经代表以前特别令人讨

厌的$\sqrt{-1}$的符号i，浮现出一个新观念：符号本身可以体现量值、方向和旋转。这就好像符号本身有某种智能。

什么是优良的数学记法？就像那些最发人深省的问题一样，这个问题的答案并非这么单纯。无论符号是什么，它必须发挥模式的启示者、一般化的指示者的功能。它本身必须具有智慧，或至少它必须支持我们自身的智慧，帮助我们思考。它必须能够指示即将到来的事物、标志崭新的想法、澄清难解的概念，协助克服不用符号而以文字表述或速记带来的混乱所造成的精神疲劳。它必须导引我们自身的智慧。

马赫又说道：

> 在代数中我们尽可能把形式上相同的数值运算一次全部解决，以便仅余一些工作待个别案例处理。使用代数和分析的记号，也就是仅进行符号运算，是因为观察到，通过手工处理所有的机械式运算，我们可以免除大脑额外的负担，用于更重要、更困难的任务。
>
> 学数学的学生常常发现很难摆脱那种不安的感觉，觉得他以他的笔体现出的科学超越他的聪明才智——伟大的欧拉坦言他往往无法摆脱这种感受。

单单一个符号便能述说整个故事。

第一次使用x^n来表示x的n次方，不是发生在某个时刻。邦贝利的$1.\underset{\smile}{2}$与笛卡儿的x^n之间，相隔了半个世纪。这个概念对我们来说或许显得清楚明了，但在乘积上用符号来标记x复制了多少个，是概念的一大推进。读者不再需要数到底有几个x，那种方式会让思考停顿，打断顺畅的阅读，阻碍可以扩展概念的任何具有联想性和类似性的广泛洞察。$x^n x^m=x^{n+m}$和（x^n）$^m=x^{nm}$，其中n和m是整数，从指标符号就几乎可以马上看出这些法则。以$x^{\frac{1}{2}}$来表示$\sqrt{x}$的想法随之而来，它产生自延伸$x^n x^m=x^{n+m}$的法则，以纳入分数，因此$x^{\frac{1}{2}}x^{\frac{1}{2}}=x^1$。

进一步思索n^x会是什么，必将引发如给定y使得例如方程式$y=10^x$，x会是什么的问题。要回答这个问题，我们会得到用加法来执行乘法运算的方法。然而，对数的发明者约翰·纳皮尔（John Napier）根本早在数学有任何符号之前很久，就已经知道答案！

符号获得了它们原本不具备的意义。但同样，符号表示有其缺点，也就是很容易忽略被表示的对象，我们继续执行运算的符号常常没有任何对象与之对应。

马赫再次说道：

> 一个计算方法的符号化表示，对于数学家而言的重要性，就如同模型或可视化的作业假设对物理学家的重要性。符号、模型和假设对被表示的事物来说，功能是相似的。但是，这样的模拟，比起原先意欲采用的符号，可以自行延伸得更远或被延伸得更远。由于这个被表示的事物与表示的设计毕竟并不相同，一边会被隐藏的东西，在另一边却是显明的。

第22章 心理学家眼里的符号

……听！横冲直撞的雪！

被太阳惊醒的雪崩！它的重量，

经过三次雪暴冲刷，聚积成庞然大物。

成千上万的思想，在逆天的心灵中，

如同片片雪花聚集，直到某些伟大的真理

被松绑，而列国同声回应，

从根开始动摇，就像此处的山岳一样。

——雪莱《解放了的普罗米修斯》

青蛙轻易就能捕捉运动中的昆虫，却无视最肥美的家蝇正停在它的面前。苍蝇可以安然地在青蛙的背上爬，无须担心被猛然吞掉。放一盘死苍蝇在青蛙前面，它会像个花园石饰品一样坐在那里。那可怜的青蛙宁可饿死，也不会攻击那些不动的东西。

我家院子的池塘里遍布各种大小的青蛙。我看到一只青蛙，但它没看到我——没有真的看到。它的眼睛没动，但如果它的身体动了，它会重新定位自己，让这个世界跟着它一起旋转。从池边草地拔下一根长芦苇，非常慢地把芦苇尖端靠近青蛙的眼睛。保持稳定，那只青蛙会这样就坐在那里，就像是盯着池塘对岸。前后摆动芦苇末端，一个舌头迅速从那只轻易被戏弄的青蛙口中吐出，抓住那根芦苇。但一只昆虫通过它不眨眼的视线范围时，它大概会快如点357玛格南手枪射出的子弹一样突然抓住那只昆虫。

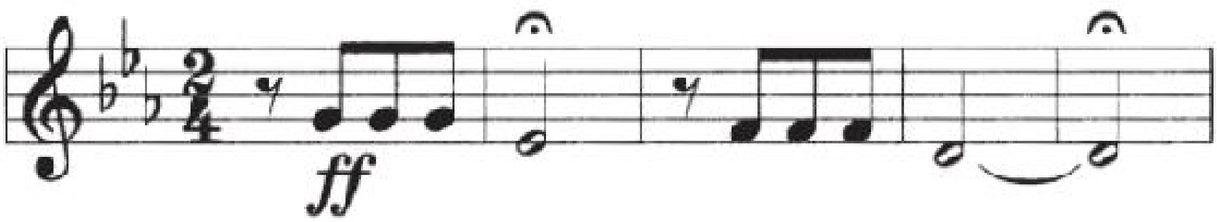
“它会失误吗？”我曾请教神经生物学家杰里·雷特温，他写了一篇影响深远的论文，讨论青蛙到底在看什么。“嗯，”他说，“只要是持续在它的视线范围内移动的东西，它都会记得，不会分心。”

青蛙看得到移动。它能够如此精准地抓住苍蝇，是因为在它的视线范围里没有其他会妨碍混淆它的东西。人类在白色或单色背景下抓

苍蝇没问题，一旦苍蝇移动到背景混杂的场景里，我们就找不到它了。

符号提供了一片空白的背景，我们可以在那个背景中仔细思考纯粹的意义。就像透过青蛙的眼睛观看一样，符号帮助我们观察分辨构成要素：从可随意使用的东西中看到不可或缺之物，从一团乱麻中看到基本原理。

你如果把方程式 $x^2-ab=0$ 放在我面前，我马上就知道 $x=\pm\sqrt{ab}$。但是，我也会看到一个正方形与一个长方形很想要互相比较。我看到一个小小的提问者在我心灵的那块白板前面扭动："与长为a且宽为b的长方形面积相同的正方形，其边长是多少？"

我认识的每一个数学家都会看到同样的这个小小提问者。这就像把下面这段乐谱拿给一位音乐家看，或就这个问题来说，给任何看得懂乐谱的人看：

他的心灵会听到一段四个音"短－短－短－长"的乐旨演奏了两次，并知道这是"贝多芬C小调第五号交响曲《命运》，作品67"的开头乐旨。

我的小小提问者会让我的脑海中浮现出几种脑成像。可能有一个是几何式的，其中两个图形，一个正方形和一个长方形，我可以重组长方形而把它变成一个正方形，以此来比较两者。由于a和b没有特定值，只能用一种符号式的操作来进行。我会采用在学校里学到的代数规则：在等号的两边加上 ab，然后 ab 开根号，得到 $x=\pm\sqrt{ab}$。

与此同时，我的头脑里可能立刻涌现出几百个特定的例子，立即搜寻这样的方程式出现的其他情况，其中a=…−3，−2，−1，0，1，2，3，4，$\sqrt{2}$，π…，b同样取值。那些特殊的情况会给我一些特定的长方形的固定意象。当a=3且b=12，将两数相乘，我会看出一个面积36平方单位、边长6单位的完美正方形。但若a=3且b=10，我会寻找一个面积30平方单位的正方形，不过有点不清楚这样的正方形边长会是多少。

此时，我的头脑必须进入一种次级模式。我必须思考如何开方，以及长年累积的关于开方的所有信息。唉！30的平方根是什么？如果我最近没有算过这个答案的话，这个问题真的很难。我会想到它小于5.5，且大于比如5.2。但接着我会告诉自己，我不是真的想知道它到底是多少，要么$\sqrt{30}$这个答案就够好了，不然$\sqrt{2 \times 3 \times 5}$也不错。

19世纪早期，德国博物学家戈特蒂尔夫·冯·舒伯特（Gotthilf von Schubert）撰写了一本极具影响力的关于梦的著作，据说该书影响了弗洛伊德和荣格。舒伯特谈到，我们以梦的视觉语言（traumbildsprache）来做梦，“一种高阶的代数”，而非用口说的语言。我们看到的图像是神话的符号和世界各地人们的仪式，但图像大多沉默不语。除了少数有言语活动的时刻，做梦的人发出的声音对任何醒着听到的人来说，都像是含糊的闷声。即使做噩梦惊声尖叫，在梦里也是无声的；痛苦的做梦者挣扎地发出最微弱的声音。

1940年，美国心理学家卡尔文·霍尔（Calvin Hall）和弗农·诺德伯（Vernon Nordby）开始收集梦。接下来三十年间，他们收集世界各地所有年龄的人的五万多个梦，汇整摘录了其中部分梦境。通过一种分类体系，他们发现这些随机从散布世界各地的不同族群收集来的梦，相似多于相异。

为什么梦的那些主题在世界各地这么多不同文化中反复出现？霍尔和诺德伯称它们为“典型梦”（typical dream）——“这些**典型梦**，我们会如此称呼它，是因为几乎每个做梦的人都做过这些梦。这

些典型梦表达了所有做梦者共同的关切、当务之急和兴趣所在。它们可以说是构成了人类心灵里那些普遍固定不变的东西。”

为什么呢？可能的答案是，图像语言早于口说语言，而梦是集体潜意识的一部分——这是荣格的理论。头脑中的图像，曾经赋予人类一种用于生存的强而有力的原型语言（protolanguage）。曾有一段时间，人类**思考**和**沟通**时所用的声音不比猎犬吠叫声更复杂。最早的口说语言很可能是以单音发出的嘟哝声或表示需求的声音。鸟不会通过对自己啭鸣来**思考**。鸟类根据可代表鸟巢的模型意象来筑巢，但不完全确定自己在做什么，但是鸟类筑巢是根据源自其中枢神经系统的本能感知的指示。它的日常行动带着一种物种特有的行为模式意义。

早在我们拥有任何今日我们会称为语言的沟通工具之前很久，我们的能力就演化到可以理解视觉意义。所以，我们会倾向于期待意象才是直觉认知最重要的部分，而非文字。我们可能无声地与自己闲谈、与自身对话，但我们所见的意象更为原始，我们不需要文字就能了解我们所看到的。我们可以用言语来解释我们的意象，但这样的解释对思维来说是不必要的。

意象和声音让看不见的思维的感官表达得以看见，让听不见的思维的感官表达得以听见。为了让感官思维发挥任何作用，必须有某种转换符码，将某种意象或声音带入意识中。意象是原始的。书写文字和数学符号则是发明出来的。在林中散步可以观察到许许多多意象——大小不一的石头、掉落的树枝、涓涓细流浸湿的叶子、绿草地、透过树梢瞥见的蓝天。这些都是无法言喻的。相反，它们成为意象，储存在大脑中凝视思想的部分，天知道那是在哪里。它们会变得混淆，然后通过与真实事件和心智意象的类似记忆做比较及联想，跟其他经验综合起来。

数学中的符号也许真的与来自感官经验的符号不同，也就是那些在梦境、神话、仪式和诗作中可清楚理解的符号，一如美国哲学家苏珊·朗格在她1967年的著作《心灵：论人类情感》（*Mind: An Essay on Human Feeling*）中所言。该书是她的最后一部著作，影响深远。然而，我们读方程式的当下——不管是简单的还是复杂的——意象（images）伴随着语言映像（verbal reflections）在头脑中形成，那些映像让人想到与以前曾看过的事物的多重隐喻连接和联想。有人可能会说，一个人的知识不过是头脑中意象与语言映像的总和。就像在林中散步一样，从集体的数学旅程中所有过去所做的符号探险里，

产生了一个综合体，一种源自抽象化过程的综合产物。朗格于1954年写道：

> ［理解符号的力量］产生自一种无意识、自发的抽象化过程，这种过程无时无刻不在人类的头脑中进行着：一种在任何经验组合中认出共通概念，并据此形成个体观念的过程。

不同于视觉观念，语言映像需要一丁点的意识来协助克服其暂时性，然后它才可以形成意义并安全地储存在长期记忆里。朗格又说道：

> 意义的指定都不是约定的，在流逝的声音之外没有永恒，但那短暂的联想就是理解乍现的瞬间。那持续的效果是，就像说话对于心智发展最初的效果一样，让事物是可想象的（conceivable），而非累积命题。

我们无意识不自觉的思想与我们有意识的思想交互作用，为我们的思考赋予意义。如果没有真实世界的感官经验让人无意识想到的难以言传的感知，这样的意义由何而来？尽管符号和文字帮助形成我们的思想和观点，但只有符号能够将可传达的想法的复杂内容，具体表现为前后一贯的表达。当然，文字也能达到同样的效果，而且在说明思想和想法时是必要的。但由于文字只能一次飞快处理一种想法，在为了完整表达那个想法而不断袭来的文字冲击下，我们会迅速陷入混乱的困境。虽然数学中的符号经过定义它们的说明文字来严格定义，它们仍可让人认识到那些文字本身未能直接表达的暗示性思想。

谈到代数时，视觉概念超越了物质世界的任何类似性。这无妨，就像我们提过的，数学的工作不是关切物质世界，也不是关切我们所谓的“实在”（reality）。符号的一致性和意义是数学的基本要素。确定性亦然，想象力亦然，创造性过程亦然，假设亦然，超越经验的信念亦然，知识的冒险亦然。而以今日的复杂性而言，做数学没有比通过符号式的拟想来进行更好的方式了。

现今，数学表达的类型各式各样。有些是图像式的（iconic），它们就像它们所表示的东西。有些真的是符号式的。还有些是纯粹用于索引性的（indexical）。美国诗人及哲学家埃米莉·葛萝舒兹在她的著作《数学与科学中的表达及有效的模糊性》（*Representation and Productive Ambiguity in Mathematics and the Sciences*）中主张：“那些随我们支配的表达以及我们结合它们的方式，决定了我

们可以如何有系统地阐述和解决问题，如何辨别项目和阐明程序，以及如何提供论据并给出说明。还有应该如何理解这些表达，它们的意涵和意义，必须在一种既有的解题传统中，适用于它们的用途。”

最近在一场于波士顿举行的数学研讨会中，我设计了一个符号认知的小实验，其中访谈了几位同事，他们都是数学教授。这项科学设计几乎称不上符合标准。在我的笔记本电脑屏幕中间是一个符号表达式，里面有一个平方根，还有一些平方。这个特定的表达式不是那么重要（参见附录C）。每次访谈一开始，我都会指着笔记本电脑的屏幕问道：“当你看到这样的东西时，你心里想到了什么？”在每一个访谈案例中，受访者都会一阵沉默，然后我会告诉受访者这没有对或错的答案。接着，他们会尝试作答，通常是与那个方程式的图形有关的某种几何论证。其中一个回答说“这可能与椭圆有些关联”，另一个回答“这是个圆锥”。

在某个时间点，一个明显的提示会以一个新的表达式的形式淡入屏幕顶端，并带着两个箭头直直指向原来的方程式。它会停留整整十秒。受访者直盯着屏幕，十秒后，方程式和箭头淡出。

我以这种方式访谈了九个人，其中只有两个人没有尝试将整个问题与问题所给的方程式图形联系起来。但显示了十秒那个淡入又淡出的奇怪方程式后，两个受试者就有了同样的想法。他们的答案正如那个提示所暗示的，是二次方程式的一般解。没有言辞显示那两个受访者注意到屏幕上那个淡入淡出的方程式。最后，我问每一位受访者，他们思考那个问题时，是否看到我的笔记本电脑屏幕上有任何不寻常的东西。他们睁大了眼睛。每一个人，包括例外的那两人，都说没看到任何东西淡入或淡出。

如果我的笔记本电脑是以文字表述而非符号来呈现那个方程式，会发生什么事？任何人都必然会将措辞转变为符号。但若我们仍生活在卡丹诺那个16世纪中叶的数学世界，那个时代的人只懂得以文字的方式表示的二次方程式解（就像婆罗门笈多甚至远在7世纪时所做的），从以言辞描述的我的淡入淡出方程式所做的提示，会这么联想到吗？

如同我的方式，问一个“当……时，你心里想到了什么”这样的问题，让我们回想到20世纪中叶社会科学实验是如何进行的，当时只有很少数的方法来测量反应。我的样本数太小，没有办法真正计算答

案出现的频率。再者，即使样本数大得多，这个实验仍必须将两个现代的观念纳入考虑，这两个观念是关于新近的联想如何与大脑实时反应结合的。我们现在知道，我们都受到启动效应（priming effect）和锚定效应（anchoring effect）两者的影响，这两个潜意识的至高力量可以操控我们有意识的推理，它们受到广泛研究。

启动效应告诉我们，我们的行动和情绪受到我们对新近事件的经验影响。举例来说，如果你被要求在“S_ _P”的空格中填字，而若你刚好洗了手，你可能会写“SOAP”（肥皂）；而若你刚好坐下吃了晚餐，可能会写“SOUP”（汤）。弗洛伊德学说也有类似的象征性关系，被要求想一个让你觉得羞愧的行动时，你会写“SOAP”，佛洛伊德的解释是，肥皂是被玷污的灵魂的清洁剂。

锚定效应不同。它无意识地将我们固定在一个小范围的联想里，倾向于将我们的看法与某种直觉的偏差紧紧固定在一起。回顾1974年时，阿莫斯·内森·特沃斯基（Amos Nathan Tversky）和丹尼尔·康纳曼（Daniel Kahneman）进行了一项实验，要受试者猜想在联合国成员中非洲成员的比例是多少。转动一个标记了数字0到100的轮盘。轮盘会在一个数那里停下——比如X。受试者首先被问到，X比问题的答案大还是小。接着，受试者被要求估计要比那个数增加或减少多少才是正确答案。结果很奇异，看到轮盘停在10的那一组，他们对联合国成员中非洲成员的比例的估计中值是25；而看到轮盘停在65的那一组，估计中值是45。正确答案是30。命运之轮怎么会与1974年联合国的成员数有关呢？

以问一个“当……时，你心里想到了什么”这样的问题来做实验时，我们应该了解，锚定和启动在实时的联想中会如何导致反应产生偏误时，扮演着重要角色。一个刚接触过某类想法的受试者，可能既被那个想法启动，也被那个想法锚定。然而，我确实认为，阅读数学符号时，启动和锚定会明确地引领至新的成果。阅读数学符号时，启动和锚定共同具建设性地引导我们通过连续不断相互竞争的联想，那些联想同时吸引某种优先的关注。锚定会把我们固定在当下的前述思想里，但阅读数学时这或许是好事。

认知心理学家基思·史坦诺维奇（Keith Stanovich）和理查德·卫斯特（Richard West）告诉我们，我们在两个层次上思考，他们将其标记为“系统1”和“系统2”，避免使他们的实验带有偏见。现在，我倾向于称呼它们为“自动模式”（auto mode）和“对焦模式”

（focus mode）。自动模式不需要努力，也不需要做判断去控制意识，而对焦模式表示要努力控制，聚焦于思考的对象。我们可以沿着空荡荡的高速公路开车听音乐，一边跟问着2＋2是多少的小孩说话。这不费吹灰之力。我们阅读比如你正在读的这本书时，同时使用了对焦模式和自动模式。我们阅读数学时，不管多么简单，我们都会用两种模式。我们使用两者，就是因为对焦模式会影响自动模式。怎么会这样？

克里斯托弗·查布利斯（Christopher Chabris）和丹尼尔·西蒙斯（Daniel Simons）的著名实验“看不见的大猩猩”（Invisible Gorilla），显示对焦模式可能会如何干扰自动模式。“不经心视盲”（inattentional blindness）这样的实验不是新发现——当注意力集中在一个目标上时，没察觉到可看见而出乎意料的物体。晚近有一些根据20世纪50年代和60年代进行的听觉研究和其视觉模拟所做的实验。查布利斯和西蒙斯的“看不见的大猩猩”实验很引人注目。他们请学生当演员，录制了一段一分钟的影片，学生分成两队运球和传球——一队穿白色球衣，另一队穿黑色球衣。受试者被要求默数穿白色球衣的球员的传球次数，忽略穿黑色球衣的球员的任何传球。影片一结束，受试者即说出他们计算的传球次数。在影片中途，一名女学生穿着全套大猩猩装走过球场，直接停在摄像机前方，捶打胸脯，然后走开。影片结束后，询问受试者一系列问题：

> 问：你在计算次数时，注意到任何不寻常的事吗？
>
> 答：没有。
>
> 问：你注意到任何不是球员的事物吗？
>
> 答：嗯，有几部电梯，还有墙上画着一些S。我不知道画那些S是什么意思。
>
> 问：你注意到任何不是球员的人吗？
>
> 答：没有。
>
> 问：你注意到一只大猩猩吗？
>
> 答：一只什么？！

大约半数受试者没有注意到那只大猩猩，一只大猩猩直接走过球场中央！大猩猩与这项任务无关，因此， 注意力不须放在那里，且因此，大猩猩是看不见的。

这项实验的目的不是要告诉我们，在我们做数学的时候，头脑如何运作；相反，它是要说明，当需要集中视觉注意力时，我们可能没看到意料之外的东西。但是，狭义地说，这确实适用于数学。我认识很多数学家，当他们深入思考一个问题时，不会察觉到房间里有只活生生的大猩猩。有多少次我不承认听过太太跟我说过某件事？大猩猩可能在房间里而我不知道，但若她碰到我正在研究的问题，她不需要捶打胸脯，便可让我知道她在那儿。

“看不见的大猩猩”实验只适用于对意外出现在视线范围内的物体的不经心视盲。数学问题**里面**的大猩猩是什么？回到我在美国数学学会（AMS）与美国数学协会（AMA）联合会议中所做的实验。两个人看到了某种东西，让他们想到之前曾看过的某种东西。这九个受试者全部在潜意识中接收到相同的提示，是什么让两个受试者发现某种代数式的东西，而七个受试者寻求图形意义上的联系？或许这只是数学头脑类型的问题，就像庞加莱（Poincaré）所说的。他写道：

> 在我们的学生当中……有些人喜欢“用分析”来处理他们的问题，其他人“用几何”。前者无法“在空间中观察”，其他人则很快厌倦冗长的计算而变得不知所措。

这里有一些答案告诉我们，为什么有些人立刻借助几何空想，其他人马上诉诸分析。当然，有许多研究尝试理解直觉的心理学和神经心理学层面，以及涉及理解数学证明、技巧或计算的创造性天赋——20世纪早期的雅克·哈达玛（Jacques Hadamard）和庞加莱，20世纪的杜宾斯基（E. Dubinsky）和波利亚（G. Pólya），接着是乔治·雷可夫（George Lakoff）、大卫·格尔瑞（David Geary）、斯坦尼斯拉斯·狄汉（Stanislas Dehaene）、大卫·拓尔，以及当今的其他学者。但是，要解答阅读一段文字式叙述对比于它的等价的符号式叙述时心里想到了什么，关于这个问题，认知神经科学的实验现在仍处于需要审慎处理的婴儿期。我现在所谈的不是以某种颅相学的方式精准指出脑中的数学思想，也不是用脑中神经心理学高速公路上的某种GPS来定位数学思考。我的问题应该不会太难回答，不过看来它成了难题。实验心理学家也不会对这个问题有多大兴趣。数学认知实验真正的难题是，人类有太多相异又富有想象力的思维模式，无法让分析具有决定性又有趣。我们动脑思考的方式多少都有些不同，个中差异很精妙，利用丰富多样的思考形态，形成和说明了人之所以为人的可贵之处。

对于这样的问题，最有趣的缜密研究来自狄汉的实验室，巴黎原子能暨替代能源委员会（Commissariat à l’énergie atomique et aux énergies alternatives， CEA）的认知神经显像小组（Cognitive Neuroimaging Unit）。狄汉和他的学生利用脑电图（electroencephalographic）技术——根据脑部在活动时会产生电流的想法——来研究思考数字与思考文字时脑部活动之间的差异。他们在计算机屏幕上快速闪现印度数字和用文字表达的数字，以找出大脑需要花多少时间来判定4小于5，精确度达毫秒（千分之一秒）。有了一些惊奇的发现。

这个实验要求志愿者在屏幕上出现比5小的数字时，用左手按下一个按钮；出现比5大的数字时，用右手按另一个按钮。脑部的活动在头皮上所产生的微小电压变化，从六十四个头皮电极片上以毫秒为单位记录下来。前100毫秒左右，记录上的电位接近零；接着，后方的头皮记录到正电位，表示枕叶的视觉区域被激活了。在这个阶段，当视觉活动进行时，狄汉发现印度数字与用英文表达的数字之间的差异是无法察觉的。但另一方面，突然间，像“four”这样的词只在左脑开始产生负电位，而像“4”这样的数字同时在左右脑产生电位。

运用心智来判定一个数比5大还是比5小这一事件——从辨认数字到按下选择按钮的动作反应——一次平均用不到半秒的时间。所以过程中发生了什么事？大约在150毫秒时，一堆专门的视觉皮质区变得活跃，推测可能是辨认数值符号的形状时未赋予任何意义。接着，在大约190毫秒时，假定这时数值的量开始编码，接近5的数字与远大于5的数字之间出现电位振幅的差异；推测起来，远大于5的数字想必比较容易辨别比5大还是比5小。

先来谈谈第一个令人惊讶之处：虽然以文字表示的数字只让左脑产生负电位，以符号表示的数字则在左右脑都产生电流，但两者所形成的电位振幅是相似的。换言之，顶下叶区（负责语言和数学运算的区域）似乎能识别数的抽象量，而不考虑记法。

第二个令人惊讶之处来自做出动作反应的第一个微小瞬间，刚好就在比较完数字准备回答之后——也就是，从文字数字或符号数字出现在屏幕上之后约250毫秒到约330毫秒的时间。在那个瞬间，左右脑的前运动皮质区与运动皮质区之间似乎产生了明显的电压差。当一位受试者先准备好用右手反应，左脑的电极片会显示负电位；准备用左手反应，右脑产生负电位。这个受试者脑中花了约四分之一秒到三分

之一秒的时间，来辨别那个文字数字或符号数字的形状，并解读它实质代表的量。

而接下来第三个令人惊讶之处是：平均而言，要再花50毫秒的时间，才能让手指肌肉收缩并实际按下按钮。即使是试着决定一个数字比5小还是比5大这种简单的工作，人们仍会犯错。出错时，额叶——与控制动作并抑制不必要行为有关的区域——立刻出现强烈的负电位，显示脑中尝试修正错误。这样的反应，发生在按下错误按钮之后不到70毫秒的惊人神速时间里，而这之所以令人惊奇，是因为它显示了对非预期反应的响应是一种心理动力的（psychodynamic）回应。

在狄汉的实验中，脑部活动的位置受到头骨往往容易扩散电位的影响而无法精确定位。要精准指出脑电活动发生的明确区域，需要更多侵入性的程序，那些程序涉及将电极片植入皮质区本身。这样的程序只有在特殊情况下才能进行，例如病人癫痫发作。1994年，耶鲁大学的特鲁特·艾利森（Truett Allison）和乔治·麦卡锡（Gregory McCarthy）以颅内电极探针进行了那些程序，结果精准定位出脑部视觉处理区的两个相邻区域：一个只对文字起反应，一个只对印度数字起反应，一个只对脸孔起反应。

狄汉、艾利森和麦卡锡不是意指脑部以某种颅相学的方式运作。他们知道即使是最简单的功能，也会激活大范围的不同大脑区域。他们知道脑部没有任何区域可以自行思考。尽管脑部没有一个单一区域可以执行甚至最简单的思考任务，但极短时间内的专门脑部活动，例如读一个词或进行一项运作，似乎确实有某种脑电活动集中的现象。狄汉认为，大脑是“迟钝代理物所形成的一个异质群体。其中每一个代理物都无法单独完成多少事情，但作为群体，它们就能各司其职来设法解决问题”。

毋庸置疑，阅读一段文字与阅读一段符号数学是有差异的。举例来说，这句“四加三与二的差再加一”，读起来跟［（3－2）＋4］＋1这句不一样。针对两者的差异这个问题，狄汉提出两个相互竞逐的问题：我们对数学式的理解是来自我们处理语言结构的能力吗？或者它与语言无关，是依靠某种视觉系统来解析数学符号串？

这个问题的答案缩小到数学是否原本是视觉-空间式的，或根本是语言式的。保有理解简单符号代数问题能力的失语患者与失智患者之

间，似乎就有差异。大脑语言区外部病变的患者，有时很难形成印度数字及其对应的等价文字的概念。

狄汉具体探究受试者在处理1＋［4－（2＋3）］这个式子时眼睛的微小行为，以便研究脑部如何表示出带有所谓巢状相嵌的数学式，以及研究有巢状结构的符号数值计算如何与解析等价语言式密切相关。

阅读文本上的句子时，在实际读那个句子之前，我们必须留意句子中的标点符号。英文中“What”开头的句子，并不一定表示是疑问句。甚至我的微软Word拼写检查，都坚持这种句子是疑问句。

就像句末的问号，处理句子中的嵌套短语时，读者的眼睛需要稍微先于认知。威廉·福克纳著作《去吧，摩西》（*Go Down Moses*）中一篇题为《熊》的短篇故事里的一段句子，让我们全心全意读到句子结束。

> 有关大荒野、大丛林的谈话：白人愚蠢得认为自己已买下了其中的任一小块地，印第安人则狠心地声称任一小块地，一直都是他们用来转售的产业；比戴斯班少校跟他所宣称的一小块地还要广大——戴斯班心里明白他“拥有”森林地着实冒犯了人类，也侵犯一切心灵共同使用土地的自然权利。比从戴斯班少校得来的老托玛斯·沙班还要古老，而沙班也自己心里有数；甚至比继承老沙班的奇克索酋长老莫度比也要古老，而老艾萨克心里也依然有数。[1]

诚然，福克纳的句子可能不是说明英文句的合宜范例，因为福克纳往往想把整个世界放到每一个句子里。但我宽恕你。这个故事里还有一些更困难得多的句子，有一段可能是美国小说中最长的句子（超过一千六百个词）。所以容许我花一点时间再读一次上面那段句子。解析福克纳的句子，需要在阅读时大幅超前目光所至的范围，可是那个句子读起来却很有道理。

一个数学式的范例如下：

$$4(3x^2+2(-3+(2-3+(2x+1)^2+1)^2)-4)$$

只有在眼睛已经扫描过这个式子，寻找嵌在最里面的有意义运算，也就是$2x+1$之后，才会开始进行运算。整个式子一般不是从左读到右，尽管可能在往右阅读的过程中已实时做了扫描。

关于人们如何理解数学式这件事，我们了解得不多。近来，出现了一些涉及功能性磁共振成像（fMRI）、脑磁图（MEG）和脑电图（EEG）的实验，以测量进行简易代数及在符号处理任务中的反应的认知时间线（cognitive timeline），也就是测量处理初等数学式的脑部位置和速度。特别是卡内基梅隆大学的杰瑞德·丹克（Jared Danker）和约翰·安德森（John Anderson）利用功能性磁共振成像及脑磁图，研究受试者求解有三个项和一个未知数的线性代数方程式时的解析速度。他们尝试将脑部两个区域中的活动分离出来：顶叶皮质（parietal cortex）和前额叶皮质（prefrontal cortex）。顶叶皮质位于额叶的后面、枕叶的上方，主要功能是整合感官信息——特别是整合视觉上感知到的空间感和导航（方向）感的数据。前额叶皮质（额叶前部）正好在运动区和前运动区的前面，功能是协调结合复杂的认知行为，从相互冲突的想法中选择采取哪一种行动，例如好或坏、同或异。

丹克和安德森利用代数符号运算任务，检验在解方程式的两步骤认知程序中大脑的两个区域，用时间序列分离出表征转换以及从记忆中提取数学知识。第一个步骤涉及顶叶皮质，此时处理视觉接收到的信息，一个意象表征做了转换。第二个步骤涉及前额叶皮质，此时它接收到转换过的表征，并将其带入数学知识的提取区，这些知识与这个问题已进行了处理的表征会相互影响。这个两步骤程序持续来回进行。举例来说，当这个刺激是解像$\frac{x}{3}+2=8$这样的方程式，就会从前额叶皮质的数学知识中提取出8－2＝6这个差；顶叶皮质转换这个方程式，将它编码为$\frac{x}{3}=6$；前额叶皮质根据这个转换，提取出3×6＝18这个乘法运算；最后，前额叶皮质转换这个方程式，将它编码为x＝18。在每一个步骤中，大脑的这两个区域都进行不同的行动，但在数学思维的提取与表征之间，有着密切的交互作用关系。

蒙纳士大学的安东尼·詹森（Anthony Jansen）、金·马里奥特（Kim Marriott）和格雷格·耶兰德（Greg Yelland）研究过有丰富经验的数学使用者如何理解代数式。他们设计了一些关于记忆的任务，检验数学语法在将代数式编码的过程中扮演的角色，并做出结论认为，经验丰富的数学使用者对于之前看过的式子，有优良语法者确认起来比不良语法者要轻松得多。他们发现，代数式的编码主要是根据超出视觉处理层次之外的程序来进行。举例来说，优良语法形式的

字符串7－x比不良语法形式的字符串如“7（x”更容易让人回想。下面这件事或许更让人惊讶：在“4－x^2）y—7”这个式子中的4－x^2，比出现在4－x^2（y—7）这个式子中的4－x^2，更容易让人回想。

蒙纳士团队后来通过检查眼睛扫描过符号的顺序，以及测量眼睛持续注视的时间，来研究经验丰富的数学使用者如何解析代数式。

阅读文本时，我们常常会在从句和句子的结尾停几毫秒。这似乎是非常自然的事，它符合我们说话的方式，一连串名词、一两个动词，连着修饰的形容词和副词，也许还有个代词，所有这些都来自将组成要素缩略纳入的一种想法。我们这样做不仅有助于口语沟通时的呼吸换气，也将句子的结构，传递为文字形式及其在句子中的相互关系的一种配置。我们可能没有想到阅读一个数学式时有相同的阅读模式，但蒙纳士团队发现，我们阅读数学式时似乎采用了某种类似的做法：在一个算术句结尾的符号，被注视的时间明显比在算术句开头或中间的符号长得多。这显示出，我们是通过代数式的语法来“读”那些式子，就像我们读自然语言的句子一样。

狄汉的实验检验了受过数学训练的受试者在计算4＋［3－（2＋1）］这种巢状算术式时，所测量到的眼部运动。眼睛会往最里面那层（2＋1）移动，且继续移到其他层，以完成计算。换言之，发现最里面那层是瞬间的事，几乎没有从左到右读那个式子，不像如果你把式子写成文字形式时我们的阅读方式。这表示受试者快速解析一开始的字符串语法，在眼睛移动到语法层系接下来的层级之前，在每一个步骤，把运算符号和括号当作线索，恢复数字等式，以便做计算。

受试者很年轻（十九岁至二十七岁），曾在法国的大学接受一般的数学教育。那些字符串都使用四个数字1到4，加上两个加号、一个减号和两对括号。复杂度分为四个层级，标记为0到3。最高层级是3，显示了如［（3－2）＋4］＋1这样一个数学上有效的项目字符串。层级2是将外层的括号对调，打乱它们外面的符号，例如“）（3－2）＋4（＋1”就是这样的字符串。层级1是将内层的括号对调，打乱它们外面的所有符号，如“＋）3－2（4＋）1（”。层级0是打乱所有的项，成为数学上没有意义的字符串，例如“4－＋）3）（＋2（1”。

狄汉的团队发现，受过数学训练的受试者对数学式的复杂度特别敏锐，而复杂度效果发生的位置是在位于典型语言区外面的一组皮质区。解析数学式最早是从视觉处理过程中开始，而那个区域非常依赖

梭形皮质（fusiform cortex），也就是涉及从形状意象中看出文字、辨识视觉世界及从视觉上辨别有良好构造的数学符号串的一个脑部区域。

看来特殊的记法组合可能有助于我们辨识数学式的结构并处理方程式——举例来说，不适当的空格可能会混淆了语法结构的空间配置，例如2＋3×4可能得出与2＋3×4不同的答案。所以，处理数学式的早期阶段或许更像处理文字的早期阶段，两者都是先确认式子的各部分来判定它们是否有效，然后才会进入语法解析阶段。

曾经有一段时间，甚至最权威的语文学家都认为，唯有凭借着语言，人才能思考。所谓语言，那些语文学家指的是文字。19世纪伟大的德国语文学家马克斯·缪勒（Max Müller）说：“我们可能感觉得到黑暗，但直到我们给黑暗一个名字之前，直到我们能够分辨黑暗是不亮的东西，或者光亮是不暗的东西之前，我们不是处在知识的状态之中，我们只是在一种被动的恍惚状态。”我们从那种恍惚状态，经过了长足的进步才走到今天。那么动物的思考又如何呢？

---

[1] 此处译文引自黎登鑫译《熊》，台北桂冠图书公司，2001年初版第二次印刷，1—2页。——译者注

第23章 符号与意向

多年前，在科德角的几个夏天，我沿着一条土路慢跑下来时，会遇到一只拴住的大型德国牧羊犬，它会对我长吠“grrrrrhoff”——就那么一声。我会非常小声地响应一声“whoof”，这是我想象中狗儿打招呼的方式。进入夏季几天后，这只狗停止了低吼，只是看着我经过。

它的脑袋里发生了什么事？它一定知道是我经过，而且我是友善的，还会发出就像它那样的声音。所以，它的脑中是什么东西让它知道那是我呢？我的意思是……它脑中的记忆库里必定有某种图像，来比对我和其他每天早上经过的事物……不是吗？在我慢跑回来时，那只德国牧羊犬从很远的地方注视我，好像等着再看到我。它没有吠叫。亚里士多德认为，若没有意象，思想不可能发生，所以或许我的犬科朋友是在把我和它小脑袋里图像思想的档案做比对。它还能怎么想呢？

维特根斯坦告诉我们：“我们为自己制作事实的图像。”对他而言，那个图像是实在（real）的一个模型。“留声机唱片、音乐思想、乐谱、声波，彼此都具有语言和世界间存在的那种图式的内在关系（pictorial internal relation）。”

当我用文字来思考，那些文字在某种程度上会进入意象，但也可能不会。听广播与看电视是有区别的。当我看电视时，我看到的是图像；当我听广播时，我制作图像。当我做数学时，阅读符号，或只是在头脑中将两数相乘，我没有真的将那些数变成图像，但我认为我有。这是怎么回事？

这就仿佛我以含混不清的符号来思考，尽管我没有真的将那些含混不清的符号变成图像。那些图像在那里，但又不在那里。当我试着想象一个三角形，我并没有三角形的一个清晰图像，只有一个隐约的模糊意象，或顶多是三角形的一个抽象表征，可能看起来不太像三角形的某个符号，但为了代表那个三角形而存在的某物。为什么那个符号不该是三角形本身？还会有什么更好的符号？

每个人都是不同的。有些人是视觉思考者，有些人是言辞思考者，还有些人可能有难以描述的思考模式。19世纪的语文学家和心理学家声称，唯有凭借语言才可能思考这件事情是“一个事实，无须论证”。我们现在懂得更多了，见到我就低吼的朋友可能是以嗅觉来思考的。

遗传学家弗朗西斯·高尔顿（Francis Galton）说过，他几乎从未靠文字来思考，而在那些罕见的时刻确实用到文字时，它们只是一些没有意义的文字，就像是“伴随着思想的一首歌的音符。它通常发生在努力工作，获得极为明确且让我满意的成果之后，我试着用语言来表达它们时，我觉得我必须先让自己置身于另一个完全不同的智慧层面”。

至于文字，法国数学家雅克·哈达玛宣称，就像高尔顿所说的，文字既未伴随着思想发生，思想也未伴随着文字发生。

> 当我真正在思考时，我坚持那些文字要彻底从我心里消失……即使读到或听到一个问题之后，在我开始仔细考虑它的那一刻，每一个字都会消失；在我完成或放弃那个研究之前，文字不会重新出现在我的意识中……而我全然赞同叔本华所写的：“思想被文字体现的时刻，它就会死亡。”

他接着说，这也是他思考代数符号时的情况。他告诉我们，他将整体的观念表现为范畴圆圈A和B，使得若A里的每一个东西都在B，那么，可以想成A是在B里。若A里没有任何一个东西在B，那么，两个圆圈是分开的；这里没有涉及任何文字，尽管可能任何关于圆圈的思考需要有一瞬间闪过文字。想象一下，试着以逻辑的方式来理解刘易斯·卡洛尔（Lewis Carroll）一段迷人的三段论法，这段文字出现在他的符号逻辑书中：

> 没有一只爱吃鱼的小猫是不能教的。
>
> 没有一只没尾巴的小猫会跟大猩猩玩。
>
> 有胡须的小猫都爱鱼。
>
> 没有一只可教的小猫有绿眼睛。
>
> 没有胡须的小猫没尾巴。

花一点功夫，我们可以制作出有数个圆圈的一张图，来解答这个三段论法。若不用那些圆圈，或没有某种高度发展的非言辞组织模式（nonverbal organizing scheme），来对头脑说明这五个句子的逻辑，要花多长时间才能得出逻辑结论，也就是“没有一只绿眼睛的小猫会跟大猩猩玩”？

更具启发性的是，哈达玛在他证明质数有无限多个的步骤中所呈现的心智图像：

> 考虑2到11之间的质数，他看到的只是“乱糟糟的一团”。
>
> 当2到11之间的质数相乘为一个乘积，也就是2×3×5×7×11，他想象到“距离那乱糟糟的一团相当遥远的一个点”。
>
> 当他在那个乘积上增加1，得到（2×3×5×7×11）＋1，他看到“比第一个点稍微远一点的第二个点”。
>
> 他看到的那个比第一个点稍微远一点的第二个点的数，若非质数，就必须包含一个质因子，而因此必须是一个比11大的质数。这时他看到的是“在那乱糟糟的一团与第一个点之间某处的一个点”。

每当我的论文指导教授迈克尔·阿廷提到数学家所谓的阿贝尔簇（abelian variety）时，他都会画一个弯弯曲曲的图案，看起来大概像，就像他的首字母缩写。它没有什么图形上的意义，当然也不像阿贝尔簇实际的样子。但直到今天，每次听到“阿贝尔簇”这几个字，我心里就会想起那个弯弯曲曲的图案。在许多方面，它正是哈达玛那个带着显而易见的含混与有用勾连的“乱糟糟的一团”。

这里有趣的是，那些距离乱糟糟的一团无论远近的代表点，都没有可分性，也没有具质数性质的要素，就像阿廷那个弯弯曲曲的图画不像代数簇。就这一点而论，它们在概念化的过程中，扮演的是引人误解和富有想象力的角色。所以，这样一个模糊的表征模式，如何有助于完全仰赖可分性（divisibility）的逻辑程序呢？虽然哈达玛的意象似乎少了因子的性质，但它提供了一个机制，可以同时检验论证的所有要素。他需要这个意象，来让那个问题看起来有特色、具结构，有一种特性。

我花了一段时间思考这个问题，因为哈达玛的答案并不让人满意。对年轻数学家的学习来说，质数有无限多个的证明，是其中一个他会很早就学到的好证明。一个真正有天赋的年轻数学家，可能不需要参考别人的证法，就能证明这个问题。我学到的第一个证明是2的平方根不是有理数，第二个证明是质数有无限多个。所以让我来回答这个问题，当**我**重读质数有无限多个这个证明时，**我的**心里在想什么。尽管我的数学头脑远远不及哈达玛那么敏锐，我的想法还是在某一方面与他极其相似。

我也看到了那乱糟糟的一团，就像那只是一堆乱七八糟的东西，还有那些远的和近的代表点。一个不同的地方是，我可能稍微更像个文字思考者，所以我使用了真正的符号，来窥见那乱糟糟的一团。你可能会说，我那一团看起来没有那么混乱。进行到最后一个步骤时，要找一个比11大的质因子，我看到在那乱糟糟的一团与第一个点之间某处，一个真实的数标记为问号。我猜想原因是——不像一个点——问号在思维上可能有除法的性质，但一个点（对我来说）不过是一个占位符号。

$\sqrt{2}$是无理数的证明有不同的思维特点。它是所谓的归谬法，方式如下：假设$\sqrt{2}$是有理数，所以$\sqrt{2}=\frac{p}{q}$，其中p和q都是整数，且q不等于0。又假设$\frac{p}{q}$为最简形式，所以p和q不会都是偶数。将最后一个等式的两边平方，可得$2=\frac{p^2}{q^2}$。两边同时乘上q^2，得到$p^2=2q^2$，因此，p是偶数。因为p是偶数，它可以写成$2s$的形式，其中s是某整数。将等式$\sqrt{2}=\frac{p}{q}$中的p用$2s$取代，使得$\sqrt{2}=\frac{2s}{q}$。将等式两边平方，得到$2=\frac{4s^2}{q^2}$。然后，化简此式得到$q^2=2s^2$，这表示q也必定是偶数。这与假设p和q不会都是偶数矛盾。QED（故得证）。

接下来谈的是我头脑中发生的事。我其实看到了这个等式$\sqrt{2}=\frac{p}{q}$，就像（我相信）任何人都会看到一样。不然，符号方程式要拿来干啥用呢？平方的运算也是符号式的，但接着我在我的文字思维中**看到**一团的许多平方数，从4，9，16，25和36开始，最后是一片模糊的更大的平方数，如100，121和144。这并非好像印度数字真的出现了，而就像我就是知道那些数存在。一旦等式被平方之后，根号消失了，而我突然看到p出现在代表所有偶数的那些点的圆圈里。事实上，p^2是偶数并没有发挥作用。我的思维跳过了从p^2是偶数推出p是偶数这个步骤。或许是因为我这辈子已经如此频繁地用过这个推论了，我根本不用去想，它已经成为我毋庸置疑的知识集合体的一部分。

当然，$\sqrt{2}$是无理数这一证明，也成为我毋庸置疑的知识集合体的一部分——我对大一新生证明过无数次了。所以为什么不把这个证明本身想象成头脑中模糊不清的一团，可以在必要时让它变得清晰？它的确是的！这就像所有其他经过充分学习和反复操作的事一样；某个位于边缘潜意识的模糊东西可能表现为一个袋子，袋子里装了与那个简单的归谬法相关或有联系的一切事物。这是一个永远可以加料的大锅，酝酿着记忆。

谁会知道一个人思考模式的内在想法或内在挣扎呢？了解自己的想法已经够难了。我们也许能知道脑的成分及其功能，知道接受功能性磁共振成像扫描的受试者听到高兴的消息时哪个部分会亮起红灯，或者知道接受正电子断层扫描（PET）的受试者必须做有风险的决定时哪个部分会亮起蓝灯。不管我们对脑和它如何运作了解多少，我们都无法更清楚知道一个个体如何思考。这真奇妙，不是吗？

我问过我的学生他们如何思考字母表、季节或月份。每个学生都有不同的模式。我的字母表是在一块中心点为M的平衡板上，我想可能跟我的姓氏有关。当我先扫描M左边的字母时，这块板子的左侧向下倾斜；一旦扫描超过M之后，板子的右侧向下倾斜。这就好像我在板子上用我的扫描把这些字母向下压。或许我就是在跷跷板上学会怎么拼我的姓氏的。但当我必须查词典或电话簿时（我是说纸本词典和电话簿），不会有意象出现。奇怪的是，在微调找到确切位置之前，我似

乎知道我的位置，而且可以直接迅速移动到我想找的词或名字大致的位置。

而月份更奇怪。地面上有个瓷砖拼出的大圆，瓷砖上标记了各个月份的名称。我正站在距离现在的月份正好六个月的瓷砖上，沿着圆的直径看向现在的月份。我所在的瓷砖位置是早六个月还是晚六个月，似乎不是那么重要。我不知道这个奇怪的月历模式是怎么来的。我的学生当中没有人曾经用这样的方式，来表达他们对月历的时间感。

从我刚刚所说的来看，会觉得我似乎是用字母和文字来思考。并非如此。

文字——如果真的存在——是头脑中的一团雾；整个思考过程是瞬间发生的，不可能自我评价自身的思考过程。文字、图像或其他任何东西，可能是这个过程的一部分，但是它们就像一首钢琴协奏曲中的个别音符——一旦奏起小节，耳朵随之聆听。

第24章 结语

> 那些语言文字，不管是书写的还是口语的，在我的思考机制里似乎不扮演任何角色。作为思考要素的那些实体似乎是某些记号，以及多少算是清晰的意象。
>
> ——爱因斯坦

我们以不清晰的图像、模糊的符号——存在，但又不存在——思考让我们得以处理日常事务的知觉和印象。在文学作品中，意识的轨迹会持续一段时间。读陀思妥耶夫斯基的《罪与罚》，直接看拉斯柯尔尼科夫挥动斧头劈了老妇人的头那一段。当我们继续往下读时，那柄斧头扮演什么角色？为什么陀思妥耶夫斯基决定那个老妇人应该被斧头劈死，而不是死于枪下或被拨火棍打死？如果使用的是另一种武器，我们的心理会做何反应？解答就在攻击头颅的手段。头颅粉碎的意涵与痛打致死大不相同。这让读者心里产生矛盾的情绪和冲突的意象：一具血淋淋的尸体与顷刻间的人道死亡。

数学中也有类似的事，我们的表达方式导致相互竞逐的感知，可能形成了我们所倾向的思考方式。我没有办法证明这一点，所以我只能提出我的想法：符号包装了隐藏在背后的潜意识所暗示的引诱，而头脑中则涌现几百个特定的例子，立即搜寻这样的方程式出现的其他情况。为什么不呢？数学文献中充满了用简单的符号建立的方程式。那些方程式本身就成为有意义的符号，提供了与一种想法强而有力的连接，也就是某无害的事物会在看似无关的领域里一次又一次出现，有时则将稍纵即逝的思想与恒常的真实世界联系起来。

几乎所有的思考都是多轨的，一轨是有意识的，其他则否。在一个轨道上，很多动作让逻辑顺畅进行。在另一个轨道上，有一切事物的潜意识记忆，那个记忆显示了与过往的连接。而当看到一些较复杂的方程式时，无数的交感神经发挥作用，能够通过潜意识的轨道传递意义，那些轨道标示着某种连接，连接到与读者曾看过的所有那些跟经验更吻合的时刻，或是富有创意的可能性的深层潜意识思想。

阅读通常既是一种认知活动，也是一种情感活动。我们可能会读到对我们的意识不具符号意义的文字和短语，但我们发现意义是在潜

意识的层次。我们不需要知道我们所读的每一个字或每一个短语，也能理解意义。文学中的意义来自联想的经验。而有时，阅读数学和物理也是如此。

数学符号与诗中的符号不同，数学符号始于数学家的刻意设计。这不会阻碍符号发挥诗中的符号那样的功能：把经验与未知连接起来，也让隐喻的想法能转而传递意义。

就像诗一样，数学中也有原型（archetype）。如果有不证自明的真理那回事，那么，或许在我们出生时，就知道了一些关于这个世界的事。

20世纪50年代，罗伯特·佛兰兹（Robert Frantz）、马克·伯恩斯坦（Marc Bornstein）、埃莉诺·吉布森（Eleanor Gibson）和其他心理学研究人员所做的婴儿实验，改变了我们关于初生婴儿如何对不同模式做出反应的看法。他们发现出生才八周的婴儿，对某些模式已经比对其他模式更感兴趣。这在当时是一项令人兴奋的发现，因为它提供给我们证据，说明从非常小的时候，我们为了建造周遭的世界，已经开始解析我们看到的一切；婴儿自动地被某些现实世界中的特定结构模式所吸引，而这些模式不断地刺激他们的神经系统。换言之，婴儿产生兴趣的行为，直接受到他所生活的世界中的结构控制。深层的结构支配了我们的行为，并连接了我们的阅读和思考、意识和潜意识的双重轨道。

柏拉图的对话录《米诺篇》（*Meno*）中提到这一点。那段对话是关于德性（virtue）是否可教的，但它使用了证明灵魂不朽的论证。为了证明这一点，苏格拉底找了一个未受过教育的奴隶男孩来询问。经过一连串问答，在没有任何人帮助的情况下，苏格拉底让这个奴隶男孩推理出关于勾股定理的事实。[1]他能够做到，据说是因为他出生前就已知道那些事实，而他可以从前世的记忆中唤回那些知识。

这很接近弗洛伊德所谓的“人类种系（human phylogeny）的集体潜意识”。当然，弗洛伊德避开了灵魂的说法。对他而言，灵魂不是灵性的，而是整个人类种族的集体潜意识——潜意识记忆代代相传，通过民间传说、宗教，以及累积起来的如何在环境不断变动的世界中生存的一般知识。这就是我们何以知道**两点间存在唯一的一条直线**，还有大毒蛇的出现可能意味着死亡、冥界、性、生殖力、疾病或治疗。大毒蛇不会出现在数学中，除了我知道的两个例子，如蛇般弯曲

的图示指出如何在坐标格子图中挑出项目。然而，那些据以建立我们的算术和几何公理的不证自明真理，来自人类种系的集体潜意识，也是承继而来的符号的根源。

符号超越了沟通媒介的角色。它们在我们的语言中无所不在，并且在将数学意象连接到意识和潜意识、熟悉和未知当中发挥了相当大的作用（虽然或许不是核心），让我们的文化／情感的倾向富有意义，所有这些都促进了这个具有创造力的过程。

两点间存在唯一的一条直线这个数学短句，相较于罗伯特·佛洛斯特（Robert Frost）的诗作《未竟之路》（*The Road Not Taken*）的第一行[2]，并没有更不像符号。它们都因来自集体潜意识中最具支配性的潜意识力量而显得与众不同。尽管来自民间传说的古典原型符号类型当中，比如大毒蛇、鸽子、狮子等等，一般不会看到数学符号，但这些数学符号还是帮助我们把未知与熟悉联系起来。在物理中，我们有麦克斯韦方程式：四个相互关联的方程式，告诉我们电场、磁场与电荷密度、电流密度之间的关系，还有它们如何随时间变化。就像任何伟大的诗作一样，麦克斯韦方程式告诉我们的远超过语言所呈现的内容。它们形成所有电力学和光学的基础，甚至引发了关于相对论和量子力学的创造性思考。

在自然语言中，日常文字描述了我们所见、所思或想象的事物。它们具有创造不熟悉的世界的力量，并且将它们带进我们的想象里。进入那些世界只需要少数特殊技巧，除了那些来自一个文化中其他人的技巧之外。要得到那些技巧，我们需要的只是身而为人的经验。

数学不同。它通常需要技巧，有时要有天赋，还往往经历经年累月的不寻常体验。我说“通常”，是因为许多非常杰出的数学家和物理学家年轻时看不出有明显的数学才能。虽然数学多半使用符号语言，解决了复杂的赘语以简化沟通，但它也利用了一种快得惊人的心智过程，解释基本要素的合理性。而就像诗一样，数学使用一种语言结构，帮助读者了解隐藏的意义，以及用言辞难以想象的东西。

数学使用的典型符号包括运算、群集、关系、常数、变量、函数、矩阵、向量，以及集合论、逻辑、数论、概率和统计所使用的符号。个别符号对数学家的创造性思考可能没有太大作用，但组合起来，它们可以通过相似、关联、等同、貌似和重复的意象，获得强有力的关联。它们甚至会创造出我们没有察觉到的想法。

每当$\sqrt{x^2 + y^2}$这个形式出现在方程式中，数学家就“知道”它表示某种度量（metric），可能是在某个坐标系统中的距离。这自然来自勾股定理，这个定理说明坐标为（a，b）的一点到坐标为（c，d）的一点的距离是$\sqrt{(c-a)^2+(d-b)^2}$。在更高维度中，这个形式可以变成$\sqrt{x^2 + y^2 + z^2}$，或是超过三个平方项和的平方根。这样的形式可能不是源自任何物理性质，但读者可以通过几何模型激发的关联性来解释它。举例来说，半径为R的圆是由坐标平面上满足方程式$R = \sqrt{x^2 + y^2}$的所有坐标（x，y）所成的集合给定的。它把某个稍纵即逝的想法与一个几何意象联系起来。尽管来自民间传说有着潜意识力量的古典原型符号类型当中，一般不会看到$\sqrt{x^2 + y^2}$，但它还是帮助我们把未知与熟悉联系起来。

有些数学符号一开始是刻意设计出来，以便把经验与未知联系起来，而且有意地通过模拟和貌似，让隐喻的想法能转而传递意义。其他符号可能意外地达到同样的效果。例如，要表示大量但有限的数值项的“和”时，我们用大写的sigma Σ，也就是希腊字母的“S”。这个符号具有有限个分开的尖角（我说三个，但这得看读者怎么数），可能源自拉丁文*summae*（“是……的和”）。而当要将无限多项相加时，我们用S形符号 ∫，这个符号平滑而弯曲，暗示着无限和。

意义和理解可能深植于关联和相似当中，这些关联和相似是通过经验得到的，并且存在于集体潜意识中。美感上吸引人的符号所产生的文化倾向，可能会影响我们的审美观，在数学中，以及诗歌和艺术领域都是如此。数学之美——证明的优雅、解说的简明、原创的巧妙、化繁为简、有意义的关联——很大一部分，归功于精巧而整齐的符号所启发的功效。

[1] 比较精确的说法是，苏格拉底让这个奴隶男孩理解如何求作一个正方形，使其面积为已知正方形的两倍。不过，这当然涉及勾股定理。——译者注

[2] 《未竟之路》第一行为“Two roads diverged in a yellow wood”（黄树林里二路分歧）。

附录A 莱布尼兹的记法

忽略一些次要的技术细节，我们可以用$y=x^2$为例，欣赏莱布尼兹的记法。我们看到变量y的值，依变量x的值变化。如果我们取某个特定的数，比如2，接着考虑下面的比：

$$\frac{x^2-2^2}{x-2}$$

对分子作因式分解后，整齐地将比化为：

$$\frac{(x-2)(x+2)}{(x-2)}$$

此时，我们发现，分母与分子有相同的因式。因此，只要x不等于2，经化简之后可知，我们一开始的那个比其实是$x+2$。但这并非只限于数字2才成立。对任意数，比如a，我们都可以用同样的方式处理。如果我们从下面的比开始：

$$\frac{x^2 - a^2}{x - a}$$

只要x不等于a，我们最后都会得到$x+a$。现在，问题来了。我们想知道当$x-a$趋近于零时，会发生什么事。之所以想知道是因为，当$x-a$趋近于零时，上述的比告诉我们：当x非常接近a时，变量x^2随x改变的情况。当然，我们不能直接令$x-a$为0，否则我们无法进行必要的消去$x-a$而得到$x+a$*的步骤*。跳过这个麻烦的方法是，观察当$x-a$趋近于零时，发生了什么事。那会迫使$x+a$趋近于$2a$。因为a是任意选择的数，a究竟是何值不重要，那意味着原始的比

$$\frac{x^2 - a^2}{x - a}$$

趋近于一个其实只依赖于a的数。它就是我们所说的a的一个函数。用符号$\frac{dy}{dx}$表示这个函数，并称之为“y对x的导函数”。

为什么$\frac{dy}{dx}$是如此优良的符号？毕竟，求导函数的最终结果，不必然得到两式之比。我们的例子所得到的是$2a$，根本不是一个比的形式。

从大多数物理现象的问题来看，你先是知道某函数的变化率，接着想了解该函数本身——举例来说，你也许知道$\frac{dy}{dx} = x$。这里不去质疑下述未证明的符号操作，你会像每一个微积分学生所学的，将

$\frac{dy}{dx}$视为分数，两边同乘dx，得到$dy = xdx$。多么方便呀！由此，我们可推导出那些奇怪的小变量dx和dy其实确实共同依循一些代数法则：如果y是x的函数，且x是t的函数，则$\frac{dy}{dx}\ \frac{dx}{dt} = \frac{dy}{dt}$。而如果$x$和$y$都是$t$的函数，则

$$\frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{dy}{dx}$$

上式为莱布尼兹另一个杰出的符号——“积分”符号——提供了登场的舞台。积分是对一个函数做运算。为了方便说明，我们再次举个例子，比如$y = x$。对y*进行*积分运算，可得一个以函数y为变化率的新函数。

借此可导出下述结果：如果两个函数相等，那么对它们积分后，恰会差一个常数。在这个例子中，所用的积分符号是$\int _\, dx$。所以，如果我们对等式$dy = xdx$的两边同时作积分，利用此积分符号会得到$\int dy = \int xdx$。等式左边可看作一个相对于变量y的变化率为1的函数，它必定是函数y本身。等式右边则是一个相对于变量x的变化率为x的函数，结果它是$\frac{x^2}{2}$。因此，$y - \frac{x^2}{2} = C$，其中C是某个常数。

附录B 牛顿的Xn流数

假设变量x的流动是均速的，并令x^n的流数为所求。同时，若量x变成$x+o$时，量x^n会变成$\overline{x+o}|^n$——亦即，利用无穷级数法可得：

$$x^n + nox^{n-1} + \frac{nn-n}{2}\ oox^{n-2} + \&c$$

两者的“增量”分别为：

$$o \text{ 与 } nox^{n-1} + \frac{nn-n}{2}\ oox^{n-2} + \&c$$

其相比（在除以o后）会相当于：

$$1 \text{ 与 } nx^{n-1} + \frac{nn-n}{2}\ ox^{n-2} + \&c$$

现在，令那些“增量”消失，则它们的最终比将变成1对nx^{n-1}之比；而因此，变量x的流数比对变量x^n的流数，一如1比对nx^{n-1}。

上述内容引自牛顿《曲线求积术》（*Tractatus de Quadratura Curvarum*），约翰·哈里斯1723年英译本。

附录C 实验

下面是一场访谈的记录，这场访谈是关于一项符号认知实验的。2012年1月4日，我在美国数学学会与美国数学协会于波士顿举办的联合会议中进行了这项实验。

我的笔记本电脑屏幕中间呈现下式：

$$\sqrt{\left(\frac{y}{2}\right)^2+z^2}+\frac{y}{2}$$

典型的访谈过程以如下方式进行：

问：当你看到这样的东西时（指着我的电脑屏幕），你心里想到了什么？

答：嗯……（一阵沉默）

问：没有对或错的答案。我只是想知道你怎么看这个东西。

答：在根式（平方根的记号）里有一个平方和，所以这可能与椭圆有些关联……不，等等……它是个圆锥，某个方向的截痕为双曲线，另一个方向为抛物线。

此时，方程式$x^2+bx+c=0$淡入屏幕顶端，并带着两个箭头，向下指着屏幕中间的那个式子。整整十秒钟的时间里，屏幕上都显示

着：

$$x^2 + bx + c = 0$$

$$\downarrow$$

$$\downarrow$$

$$\overbrace{\sqrt{\left(\frac{y}{2}\right)^2 + z^2} + \frac{y}{2}}$$

受试者直盯着屏幕，十秒后，方程式和箭头淡出。

我以这种方式与九个人进行了访谈，其中只有两个人没有尝试将整个问题与问题所给的方程式图形联系起来。但显示了十秒那个淡入又淡出的奇怪方程式后，两个受试者就有了同样的想法。下面是其中一个受试者的访谈记录，另一个受试者的访谈记录事实上也差不多。

答：停停，也许x，y，z不是变量。

没有言辞显示他们注意到页面上那个淡入淡出的方程式，但在例外那两人的访谈中，受试者都在板子上写下：

$$\sqrt{\left(\frac{b}{2}\right)^2+c^2}+\frac{b}{2}$$

第二个受试者仿佛知道第一个受试者做过的事。

问：嗯，所以你现在看到了什么？

答：看到某种像……二次方程式的一般解吗？它是二次方程式的正根吗？

问：你想到了什么？

受试者写下$x^2+bx+c=0$。

答：不，不……（重写方程式，把$+b$改成$-b$。）

再重写方程式，把$+c$改成$-c$，受试者持续注视着新方程式，而我在过程中不发一语。过了一小段时间后，他写下$x^2-bx-c^2=0$。他自信地继续改写，最后写下$x^2-yx-z^2=0$。

问：很好！

答：（睁大了眼睛）哦……这（指着电脑屏幕中间那个式子）是方程式$x^2-yx-z^2=0$中x的正值。

我设计这个问题时，用z^2取代原先的z，使得问题比我原先预定的难度高。我的用意是让根号中的项能化成椭圆的形式，把一切变复杂。我大可以直接写

$$\sqrt{\left(\frac{b}{2}\right)^2+c^2}+\frac{b}{2}$$

而不是

$$\sqrt{\left(\frac{y}{2}\right)^2+z^2}+\frac{y}{2}$$

但我觉得这样是送分题。

最后，我问每一位受试者，他思考那个问题时，是否看到我的电脑屏幕上有任何不寻常的东西。每一个人，包括例外的那两人，都说没看到任何东西淡入或淡出。

有两个人注意到根号里的

$$\left(\frac{b}{2}\right)^2+\text{某物}$$

这让他们想到b^2-4c这个形式，求解形如$x^2+bx+c=0$的方程式时，这个式子总是出现在根号里。从图形来看，形式

$$\sqrt{\left(\frac{b}{2}\right)^2+c^2}$$

也会暗示一个正椭圆锥。这九个受试者全部在潜意识中接收到原式与$x^2+bx+c=0$之间的关联的暗示，是什么让两个受试者发现

$$\sqrt{\left(\frac{y}{2}\right)^2+z^2}+\frac{y}{2}$$

如同二次方程式的解，而七个受试者寻求图形意义上的关联？

我无法以文字告诉你，我究竟是如何知道

$$\sqrt{\left(\frac{y}{2}\right)^2 + z^2} + \frac{y}{2}$$

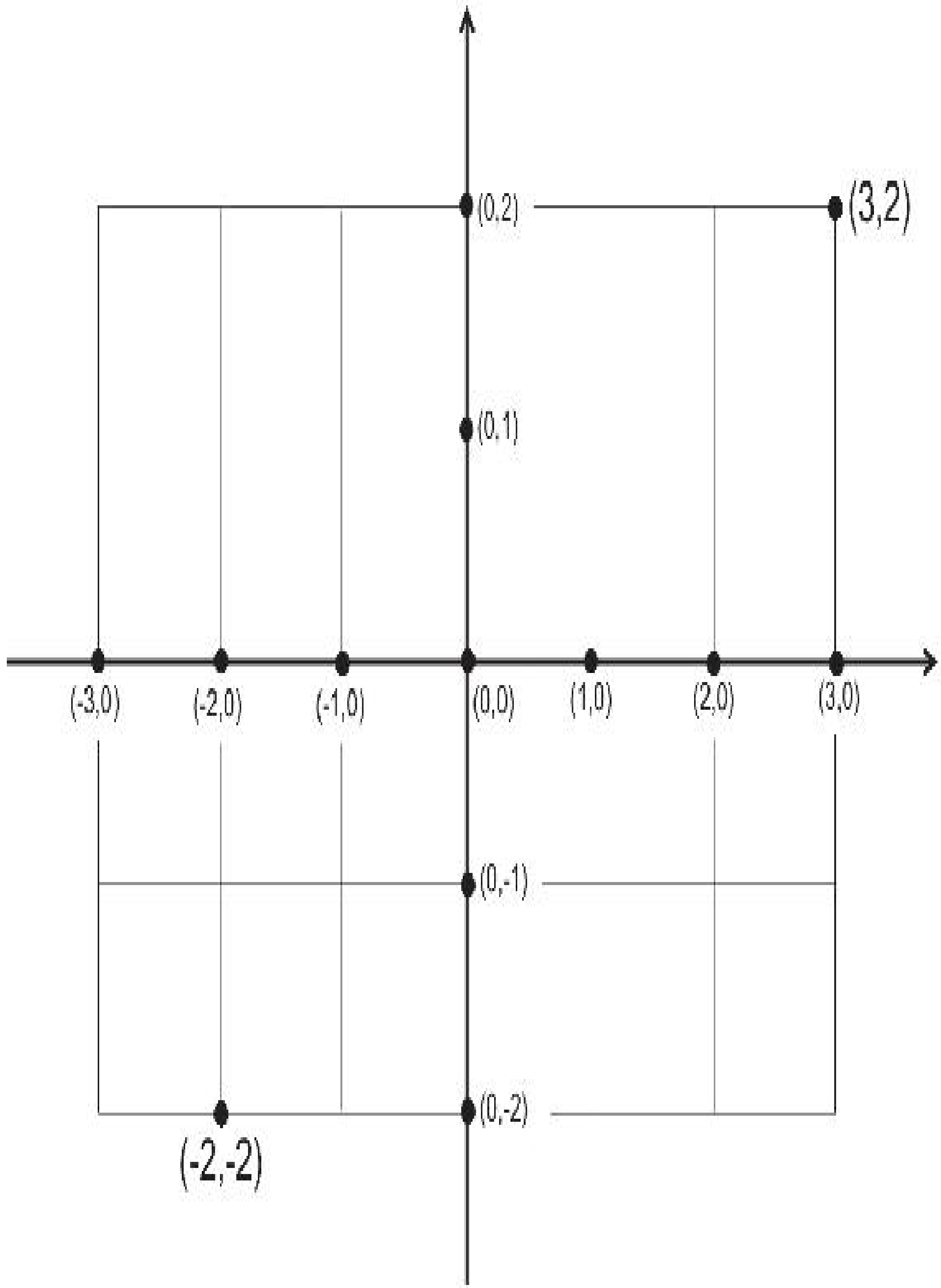
是二次方程式$x^2+bx+c=0$的正根。我只能告诉你，每当我看到根号中出现任何下述形式

$$\left(\frac{y}{2}\right)^2 + 4 \times \text{某物}$$

都会让我想到b^2-4c这个形式，从而想到$x^2+bx+c=0$的根。

附录D 将复数可视化

将复数$a+ib$表示为坐标平面上的一点（a，b）。如此一来，所有碰巧为实数的复数便会落在通过原点（0，0）的水平线上，而所有碰巧为纯虚数的复数则会落在通过原点（0，0）的铅直线上（参见图D-1）。


(0,2)
(3,2)
(0,1)
(-3,0)
(-2,0)
(-1,0)
(0,0)
(1,0)
(2,0)
(3,0)
(0,-1)
(0,-2)
(-2,-2)

图D-1 将复数可视化

每个复数都可表示成一个数对，并将其画在这个坐标平面上。但为什么我们称呼这些形如“数对”的东西为数呢？答案是因为它们遵守数的算术法则。将任意两个复数相加，会得到第三个复数：定义（a，b）+（c，d）为（$a+c$，$b+d$）。［请注意，（$a+c$，$b+d$）即为（a，ib）+（c，id）相加所得的（$a+c$）+i（$b+d$）。］那乘法呢？我们定义（a，b）乘以（c，d）得（$ac-bd$， $ad+bc$）。［请注意，（$ac-bd$， $ad+bc$）即为（$a+ib$）（$c+id$）相乘所得的（$ac-bd$）+i（$ad+bc$）。］利用这些加法和乘法的定义，所有复数的算术法则都不会产生矛盾的情况。但当我们更进一步思考时，一些有趣的事情发生了。复数的乘法有几何上的意义。乘上复数i的几何意义为逆时针旋转90度；乘上复数di的几何意义为逆时针旋转90度后，再伸缩d倍。

若使用记法$\sqrt{-1}$来表示复数i（实际上，它们的意义相同），上述讨论的一切结果依旧成立。但复数i可以从开根号的概念中，分离出旋转的意义，让头脑意识到代数结果与数的概念的延拓之间的区别。

附录E 四元数

哈密尔顿领悟到的东西，是他的四元数$x+iy+jz+kw$中的每一项，皆与其他项独立，并满足乘法法则$i^2=j^2=k^2=ijk=-1$，而且它可以表示成四元数组（x，y，z，w），并遵守交换律之外的所有代数乘法法则。他会满意如下事实：此四元数组满足$ij=k$及$ji=-k$。他也会接受-1存在超过两个平方根，事实上有无限多个！他会接受二次方程式可以有两个以上的解。这些都是要将数推广至复数以外更高维度的数系时，所必须付出的代价。这个新数系包含了复四元数，并且内嵌了三维的虚数系$iy+jz+kw$。

那么，在三维空间中，乘上i，j，k各代表了什么样的几何意义呢？旋转，我们期望它与旋转有关。但怎么转呢？如果i，j，k分别代表三维空间中，三个彼此垂直坐标轴上的正向单位向量，则以j乘上i时，会将整个三维坐标系旋转90度，并将i轴送到j轴，且使得k轴维持固定不动。它告诉我们，三维空间有两个不同的赋向模式，而物理学家必须决定使用哪一种当成约定常用的方式。换言之，木螺丝的螺纹应该设计成顺时针转入木头还是逆时针转入？这里的选择是随意的，但约定俗成的方向是顺时针。如果你大学时学过物理，可能会记得这些旋转的方向就如同空间赋向中的右手定则一样，对物理学和数学而言，它都是基础而重要的约定。

与复数不同的是，四元数在我们熟知的空间中并无适当的表示方法，而且无法可视化。对我们当中未受过四维空间思考训练的人而言，它们不存在可接受的可视化表示方法。然而，我们还是在数系的延拓过程中，将四元数纳为有效的数。它们并非以几何的方式出现，它们可能是通过符号表示的方式现形。现在，它们总出现在人们最意想不到之处。要不是欧拉在1777年提供给圣彼得堡科学院的回忆录里将$\sqrt{-1}$表示成i，要不是在他去世后于1794年将$\sqrt{-1}$印制出版

成i，要不是高斯在1801年后一贯使用i，四元数在数学史上不会这么快被发现，而给数学物理的发展带来如此重要的贡献。

致谢

要不是许多人的支持，本书不可能完成。首先，我的太太珍妮弗数次阅读草稿全文，提供编辑、结构和实际方面的建议。她是我永远的灵感来源和支柱。

我非常感激伯利亚斯科基金会提供资助，让我在意大利伯利亚斯科一如其名的天堂湾宏伟的黛皮尼别墅完成本书。这项资助慷慨地提供给我从容的时间、奢华舒适的居所，以及午后在温暖的地中海游泳，还有惠我良多的同人，包括鲁比·布隆戴尔、约翰·伊顿、路易莎·科斯塔·戈麦斯、珍妮·玛丽·刘泰、珍妮弗·萨克斯和威拉德·施皮格尔曼，他们直接或间接帮助我完成本书的最后稿本。

特别感谢审阅本书全部文稿的耐心读者：米歇尔·鲍尔、朱丽安·费恩霍尔茨和玛格丽·塞尼察尔。谢谢其他读过个别章节（通常是好几章）的人：鲁比·阿瑞安赫德、史蒂夫·巴特森、肯尼斯·布利兹、查尔斯·伯内特、巴瑞·塞普瑞、大卫·考克斯、罗伯特·道森、菲洛森·蒂雅库、拉法埃拉·弗兰奇、费尔南多·古弗、艾米莉·格罗霍尔茨、菲尔·霍姆斯、因斯·霍伊鲁普、盖泽姆·克劳利、米哈伊尔·卡茨、凯瑟琳·马祖尔-杰夫瑞斯、巴瑞·马祖尔、彼得·梅雷迪斯、金姆·普罗夫科和西沃恩·罗伯茨。另感谢我的专家顾问，他们在书信往来和交谈中提供了丰富的信息与激励：斯坦尼斯拉斯·德阿纳、大卫·吉尔里、丹尼尔·卡内曼、乔治·拉考夫、斯蒂芬·平克、伊恩·斯图尔特、大卫·塔尔和伊丽莎白·乌利维。

我自认是浸淫在一群博学多闻者当中的数学撰稿者。因此，我感谢致力研究符号史，以揭露已佚失数百年的智识成就的真正学者。克里索马里斯（Stephen Chrisomalis）在他的博士论文《数字记法比较史》（*The Comparative History of Numerical Notation*，2003）中，列举了一千零四十七件关于数字与数字系统、数字记法和记数法的学术出版品，对我的研究帮助良多。感谢梵蒂冈图书馆的保罗·维安、美国克雷数学研究所、古腾堡计划、开放图书馆、芝加哥大学数

字保存收藏、欧洲文化遗产在线（ECHO）、纽约公共图书馆数字画廊、自由基金会、科学哲学存储库、梵蒂冈图书馆、佛罗伦萨威尼斯国家档案馆、佛罗伦萨中央国家图书馆、帕维亚大学等。法国国家图书馆在线善本图书馆、网站Scribd.com、网站Ancientlibrary.com、珀尔修斯数字图书馆、恩尼奥·德·乔治数学研究中心（参考邦贝利的《代数学》文本），以及高等师范大学图书馆，让我可以居家研究，十年前得耗时经年周游世界各地图书馆的善本室才能进行这些研究。感谢乔纳森·本内特，他经营一个数学家和哲学家书信翻译的网站（www.earlymoderntexts.com）。还要感谢不再与我们同在的真正学者，他们致力于研究符号史，以揭露已佚失数百年的智识成就：弗洛里安·卡裘利、托马斯·希斯爵士、路易斯·查尔斯·卡宾斯基和卡尔·门宁格。非常感谢马立克，本书的捷克版译者，他在本书制作最后阶段发现一堆英文排版问题。

特别感谢我的编辑维奇·科恩斯的坚定支持，还有我的经纪人安德鲁·斯图尔特在我非常简短的提案中看出这项出版计划的潜力。

感谢马祖尔·杰夫瑞斯家族：凯瑟琳、汤姆、索菲亚、伊莲娜和内德。感谢马歇尔家族：达米娜、斯科特和莉娜。感谢我的兄长巴瑞及兄嫂弟妹卡罗尔·约菲和格雷琴·马祖尔，向他们永远的鼓励致意。

关于作者

约瑟夫·马祖尔是美国马尔波罗学院荣誉数学教授，他教授的课程涵盖与数学相关的领域，包括数学史和数学哲学。

同时，他也是一位著名的科普作家，他在本书中叙述了符号系统发展曲折离奇的历程，检视过去两百年间史学家对数系起源的争论，详尽细究各个文化中关于数和符号的基本原理，探究了这些符号如何透过“相似”“结合”“恒等”“类似”“言外意象”来影响我们，如何借由潜意识的结合衍生出新概念，如何在经验与未知之间建立联结，还有它们如何象征了人类的抽象知识提升到一个完全不同的层次。

关于译者

洪万生，美国纽约市立大学博士，主修数学史、科学史，辅修数学哲学、科学哲学。

洪赞天，美国纽约州立大学水牛城分校英文系毕业，现为专职翻译。

英家铭，中国台湾师范大学博士，现任台北医学大学人文暨社会科学院助理教授。

黄俊玮，中国台湾师范大学数学研究所博士，主修数学史与数学教育，主要研究领域为日本江户时期数学史（和算史）。

黄美伦，中国台湾师范大学数学系硕士，主修数学史与数学教育，研究领域为台湾数学教育史。

郑宜瑾，中国台湾师范大学数学系硕士，主修数学史与数学教育，研究领域为数学科普作品。

《人类符号简史》极富洞察力地纵观了人类历史与心智上的革命，这些革命创造了人类历史上最有用的符号　　数学符号。

《大脑与阅读》作者、法国科学院院士及梵蒂冈科学院院士
斯坦尼斯拉斯·迪昂

本书追溯了造就今日符号形态的研究历史和思考方式的曲折过程……作者马祖尔以精雕的笔法探究这个主题，使《人类符号简史》为读者提供了兴味盎然的阅读经验。

——美国《科学》杂志

马祖尔为读者娓娓道来数学符号背后令人着迷的发展历程。我们天天使用这些符号，视其为理所当然，但实际上数学语法把“数”变成“叙述”，这对门外汉来说是一种神秘又难以理解的密码。马祖尔通过他生动易读的文字将枯燥难懂的历史转变成趣味横生且内容丰富的故事。

——美国《出版人周刊》

接力出版社公众号

陈列类别◇社会科学
ISBN 978-7-5448-5222-7
定价：42.00元

Table of Contents