

ON INTELLIGENCE

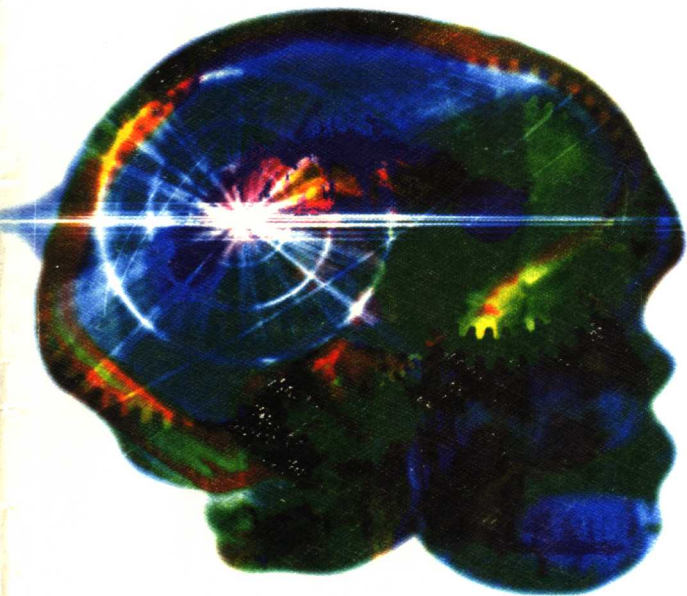
真正认识人类大脑是开发智能机器的必由之路！

—— 杰夫·霍金斯

人工智能的未来

[美] 杰夫·霍金斯 (Jeff Hawkins) 著
桑德拉·布拉克斯莉 (Sandra Blakeslee)

贺俊杰 李若子 杨倩 译



陕西科学技术出版社

人工智能 的 未来

杰夫·霍金斯

(Jeff Hawkins)

[美]

桑德拉·布拉克斯莉

(Sandra Blakeslee)

著

贺俊杰 李若子 杨倩 译

陕西科学技术出版社

图书在版编目(CIP)数据

人工智能的未来/(美)霍金斯(Hawkins, J.), (美)布拉克斯莉(Blakeslee, S.)著;贺俊杰,李若子,杨倩译. —西安:陕西科学技术出版社,2006.1

书名原文:On Intelligence

ISBN 7-5369-4018-1

I.人... II.①霍...②布...③贺...④李...⑤杨... III.人工智能
IV.TP18

中国版本图书馆 CIP 数据核字(2005)第 116759 号

ON INTELLIGENCE by Jeff Hawkins with Sandra Blakeslee

Copyright © 2004 by Jeff Hawkins and Sandra Blakeslee

Published by arrangement with author c/o Levine Greenberg Literary Agency, Inc.
through Bardon-Chinese Media Agency.

Simplified Chinese translation copyright © 2006 by Shaanxi Science & Technology Press

ALL RIGHTS RESERVED

陕西科学技术出版社享有本书中国大陆地区简体中文版专有版权,该权利受法律保护,侵权必究。

人工智能的未来

- 出版者** 陕西科学技术出版社
西安北大街 131 号 邮编 710003
电话(029)87211894 传真(029)87218236
<http://www.snstp.com>
- 发行者** 陕西科学技术出版社
电话(029)87212206 87260001
- 印刷** 西安信达雅印务有限公司
- 规格** 880mm × 1230mm 32 开本
- 印张** 8.5
- 字数** 185 千字
- 版次** 2006 年 1 月第 1 版
2006 年 1 月第 1 次印刷
- 定价** 18.50 元
-

(如有印装质量问题,请与我社发行部联系调换)

► 序 言 ◀

这本书的创作激情源自于我所钟情的两样东西——计算机和大脑。

25年来我在移动计算机技术领域孜孜以求，毫无倦怠。在美国硅谷的高科技世界里，我可谓数得上的人物——曾创办过 Palm Computing 和 HandSpring 两家 PDA(个人数字助理或掌上电脑)公司；作为总设计师，也设计出许多掌上型计算机和智能电话，如 PalmPilot 和 Treo 等产品。

然而，我的第二种爱好不但早于我对计算机的兴趣，而且对它感情更为炽烈。我对大脑研究非常着迷。我想理解大脑是如何工作的，不仅仅从哲学的角度、从笼统的观点去理解，而是从细微处去探索，扎扎实实、精益求精地去理解。我不但渴望了解智能是什么、大脑又是如何工作的，而且更期望制造出像大脑那样工作的机器。创造出真正具有智慧的机器是我的最终目标。

当今，在地球上最后要攀登的科学高峰是解决智能的问题。大部分重大的科学论题所涉及的事件往往是极端微小、极端庞大，或者涉及亿万年前极为遥远的事件。然而，有关大脑的问题却不是这样。人人都有一颗大脑，你的大脑是属于你自己的。你为什么会这样感觉，而不是那样？你如何认知世界？为什么你会犯错？你怎样才能有创意？为什么音乐和艺术可以激发

灵感?如果想知道这些问题的答案,你必须了解大脑。另外,一个关于智能与大脑功能的成功理论不仅可以为治疗脑部疾病服务,还会带来巨大的社会效益。我们要建造真正的智能机器绝不同于科幻作品中所描述的机器人,相反,这种智能机器会随着一套有关智能本质的新理论大显身手。它们会帮助我们加速了解世界,辅佐人类探索宇宙,让世界变得更加太平。

我们有幸诞生在这样的时代,一个可以解答智能问题的时代。我们这一代人不但拥有百年以来收集到的堆积如山的关于大脑的数据,而且数据收集的速度也在日益加快。单在美国就有成千上万的神科学家。可惜至今也没有形成一套有关智能和大脑工作原理的完整理论。大部分神经生物学家不大考虑有关大脑的理论,因为他们只贯注于所做的有关大脑的诸多子系统的实验,只埋头于收集更多的数据。尽管一批又一批的计算机程序师曾经尝试让计算机拥有智能,却一次次以失败而收场。我相信他们会屡战屡败,除非这些程序师不再忽视计算机与大脑之间的差别。

智能究竟是什么?为什么人脑有智能,而电脑没有?为什么一个6岁的小孩子可以在河床的石头上跳来跳去,姿态优美,而当今最先进的机器人却像行动迟缓的傻瓜?为什么3岁小孩已经粗通语言,而计算机却言语不通,枉费了半个世纪以来程序师艰苦卓绝的努力?为什么你能在1秒钟之内分辨猫与狗,而超级计算机却不能?这些都是亟待破解的斯芬克斯之谜。目前线索不少,但真正需要的是一些关键性的见解。

也许你很奇怪为什么一个计算机设计师要写一本关于大脑的书。换句话说,如果我对大脑的研究如此痴迷,为什么不将

脑科学或人工智能作为自己的事业呢?答案是,我尝试过,而且不止一次,但是我拒绝像前人那样研究智能问题。我相信解决问题的最好方法,就是利用人脑生物学提供的细节作为约束与指引,把智能当作一个计算性质的问题——这是一种介乎生物学与计算机科学之间的立场。许多生物学家基本上不接受或忽视以计算机术语的方式去认识大脑,而计算机科学家通常不相信生物学对他们来说有何借鉴之处。科学界比商界更不愿意冒险。在高科技行业,如果一个人从事新概念的研发工作,只要有充分的理由,无论成功与否,都能促进自己事业的发展。许多成功的企业家都是在早期失败之后才成功的。然而在学术界,如果新概念的研发工作经过几年时间仍没有成果,新生的事业就要夭折。所以,我决定同时追求生命中的两种爱好,并坚信商业上的成功会帮助我取得人脑研究的成功。我需要财力来支持我所追求的科学研究,我也需要学会推动世界变革的进程和推销新概念的方法,所有这些我在硅谷的工作中受益匪浅。

2002年8月,我创办了一家研究中心,取名为“红杉神经科学院”(Redwood Neuroscience Institute),简称RNI,专门进行脑理论的研究。世界上有很多神经科学中心,但唯此一家是专门从事全面的新大脑皮层理论研究,而新大脑皮层正是人脑中负责智能的部分。这就是我们在RNI的全部项目。

从很多方面来说,RNI就像一家初创公司。我们正在追逐有些人认为遥不可及的梦想,幸而这里人才济济,我们的努力已初结硕果。

这本书所表述的内容可谓雄心勃勃，它描述有关人脑如何工作的全面理论：智能是什么？人脑又是如何创造智能的？我提出的理论并不是全新的。你将读到的许多概念早已零散地存在，但将它们像珍珠般一颗颗串起来，构成一串光彩夺目的项链，本书却是首次。这也不足为奇。据说新概念往往是旧概念的重新包装和重新解释。这句话的确适用于这里提出的理论。然而包装和解释能产生天壤之别的效果，即仅有大量细节的素材与一个能够令人满意的理论之间的差别。我希望这个理论会打动你，使你产生类似的感觉。我听到的一种典型的反应是：“有道理。我本来永远都不会这样看智能，现在既然你说了，我倒是能理解了。”一旦拥有这些知识，大部分人都开始用不同的眼光打量自己。你开始观察自己的行为，说：“我明白刚刚在我脑子里发生的一切。”希望读完这本书的时候，为何想其所想，又为何行其所行，你会对此有一个新的认识。也希望有些读者由此得到启发，根据书中勾画的原理，专注于建造智能机器的事业。

我常常把这个理论和我研究智能的思路称作“真智能”（real intelligence），以示与“人工智能”的区别。人工智能科学家试图通过编写程序，让计算机模仿人类的行为，却没有回答智能是什么，其含义又是什么。他们遗漏了建造智能机器最重要的部分，即智能！“真智能”认为，试图建造智能机器之前，我们必须理解人脑是如何思考的，而这里并没有丝毫人工的东西。只有到此时，我们才可以考虑怎样建造智能机器。

本书前五章首先介绍为什么以前在理解智能与建造智能机器方面的努力都失败了，然后提出并进一步论述我称为记忆-预测架构(memory-prediction framework)理论的核心概念。第六章则细说人脑如何实现记忆-预测模型，换句话说，就是人脑究竟如何工作。接下来讨论这个理论对社会及其他方面的意义，对许多读者来说，这或许是书中最引人深思的章节。最后一章则讨论智能机器的建造与未来。但愿此书能令你神往。以下是我们将会一一探讨的问题：

计算机可以拥有智能吗？

数十年来，人工智能领域的科学家宣称，当计算机足够强大时，就可以拥有智能。我不这样认为，而且我会解释为什么。人脑和计算机的工作原理完全不同。

神经网络方面的研究会不会有助于智能机器的制造？

大脑固然由神经元组成的网络所构成，但若不首先理解大脑的工作原理，在创造智能机器方面仅凭简单的神经网络研究，将不会取得比开发计算机程序更大的成功。

破解人脑工作原理之谜为何如此困难？

多数科学家认为大脑是如此复杂而高深，我们需要花很长的时间才可以理解它。我不同意。复杂而高深是思想混乱的征兆，并不是其原因。反之，一些直觉的假设和错误的假设把我们误导了，其中最大的错误就是相信智能是以智能行为来定义的。

智能若不是以行为来定义，又该以什么来定义？

大脑使用大量的记忆来创造世界的模型。所有你知道的和学到的事物都储存在这个模型里。大脑用这个以记忆为基础的模型去不断地预测未来的事件。预测未来的能力才是智能的关键。我将深入地描述大脑的预测能力，它正是本书的核心概念。

人脑是怎样工作的？

智能产生于新大脑皮层。新大脑皮层尽管多才多艺而且适应力极强，然而它的结构细节却出奇地规则。新大脑皮层的不同部位，不管是负责视觉、听觉、触觉，还是语言的部分，都遵循相同的原理工作。理解新大脑皮层的关键在于理解这些共同的原理，特别是它的层级结构。我们将深入分析新大脑皮层，展示它的结构是怎样捕捉世界的。这些讨论将是全书中最具技术性的部分，但只要你对此感兴趣，即使你不是科学家也可以读懂。

这个智能理论的含义是什么？

有关大脑的理论可以帮助解释许多事物，例如我们怎样才有创造力，我们为什么会觉得自己有意识，我们为什么会持有偏见，我们是如何学习的，为什么有时却“老狗难学新把戏”，等等。我将会讨论许多这样的话题。总的来说，这个理论能帮助我们洞察自己是谁，又为何行其所行。

我们能够建造智能机器吗?它们又将会做什么?

是的,我们能,而且我们会。今后数十年,我预见这种机器的性能将会朝着有趣的方向迅速地演化。有人害怕智能机器会危害人类,但是我强烈地反对这种危言耸听的想法。人类是不会被机器人打垮的。建造在物理、数学等高层思维方面超越我们的机器,比建造科幻小说中会走、会说的机器人要容易得多。这种令人难以置信的技术正是我要探索的。

我的目标就是要以人人能懂的方式说明这个新的智能理论和人脑的工作原理。一个好的理论应该容易理解,不应该隐匿在专业术语或者纷繁复杂的论据之中。我将从一个基本架构出发,然后逐步添加细节,有的单纯是逻辑推论,有的则涉及人脑连接回路的特定方面。我所提议的某些细节肯定会有错,不过这在任何科学领域中都是难免的事。要发展一个完全成熟的理论需要多年,但这无损理论核心概念的力量。

多年前,当我最初对大脑发生兴趣的时候,我去当地的图书馆想找一本解释人脑如何工作的好书,那里有几乎所有有趣的主题,包括相对论、黑洞、魔术、数学等,这些都是当时令我神往的东西。十几岁时我早就习惯了找到有关的好书,然而却寻不到一本令人满意的说明大脑的好书。我终于认识到,没有人知道人脑实际上是如何工作的,甚至连一个基础的或者未经证实的理论都没有。这是不正常的。例如,当时虽然没有人

知道恐龙是怎样灭绝的，但与之相关的理论却很多，而且全部都可以读到。关于人脑的情况则完全不一样。起先我不相信，我们竟然不知道这个关键的器官是怎样工作的。在对已知的有关大脑的研究过程中，我树立起一种信念：一定有个简明直接的解释。大脑不是魔术，而且我觉得答案甚至不会像魔术那么复杂。数学家保罗·厄多斯(Paul Erdos)相信，最简单的数学证明早已经写在“天书”里，数学家的工作就是解读天书，找出这些证明。同样地，我觉得智能的解释就在“天书”里。我能体验个中滋味。我想读懂那部“天书”。

25年来，我一直都有一个梦想，希望写出一本解释大脑的简明系统的小册子。它好像吊在我面前的一根胡萝卜，不断激励我。这个梦想成就了正握在你手中的这本书。我向来不喜欢复杂，无论是科学还是技术。这种偏好反映在我所设计的产品中，它们往往以容易使用而知名。最有力量的东西往往都是最简单的。这本书提出一个简明直接的智能理论。我想你会喜欢上这本书。

▶ 目 录 ◀

序 言

第一章 人工智能 (1)

有些科学家认为计算机可以做到人脑所能做的一切,人工智能机器的制造只是编写程序的问题。于是,一代又一代科学家们不断尝试,从 Eliza 程序到“积木世界”,再到著名的“中文屋”实验。我们清楚地看到,尽管这些程序可以模拟大脑,但它们不可能具有智能。直觉告诉我,这样的传统方式不可能制造出真正的智能机器。只有研究大脑、认识智能才是最终的解决之道。

第二章 神经网络 (17)

20 世纪 60 年代后期,一种新的智能机器研究途径崭露头角,它就是神经网络。相对于人工智能而言,这是一个进步,因为它是以真正的神经系统为基础的。但令人遗憾的是,神经网络的发展很快就固定于一些简单的模型之上而停滞不前。它对于时间和反馈的忽略使我进一步坚信,要制造出智能机器,必须从天然的智能引擎——新大脑皮层开始探索。

第三章 人脑 (35)

从各个角度观察人脑,其外表几乎没有什么区别,但人脑的内部包含许多不同的功能区,这些功能区在一个复杂的层级结构中相互联系。此外,神经科学家蒙卡斯尔认为,大脑皮层不仅在外表和结构上极其相似,进入大脑皮层的输入信息也是基本相同的,不论是听觉、视觉还是触觉。那么,我们的大脑到底是如何感知这个世界的呢?

第四章 记忆 (61)

大脑皮层和计算机完全不同,它不会计算问题的答案,而是使用存储的记忆来解决问题,做出反应。由此,我们又接触到一些崭新的重要概念,如存储模式序列、自-联想记忆、恒定表征等等,它们发生在大脑皮层的每一个区域,无处不在。

第五章 智能理论新架构 (83)

我们的大脑是用一种平行的方式连续不断地对我们所在环境的构架进行预测。这是大脑皮层的主要功能,也是智能的基础和必要因素。那么,我们的大脑皮层是如何进行预测的?

第六章 大脑皮层工作原理 (105)

大脑皮层为了预测,必须能够记住并且存储关于事件序列的知识。要预测新的事件,它又必须形成恒定表征,而恒定表征存在于包括视觉、听觉、触觉和联合区等大脑的所有区域。大脑皮层区域到底是什么样子?皮层区域的详细工作原理是怎样的?为什么需要反馈呢?有了反馈就能解决问题了吗?大脑皮层是如何学习的?

除了记忆、反馈和恒定表征以外,大脑还需要按照世界的本来面貌建立并存储一个世界的模型。大脑存储世界的方式就是形成新的分类和新的序列。

第七章 意识和创造力 (181)

创造力来自哪里?意识是什么?想像力又是什么?动物是否有智力?人的智力与动物的智力有何区别?什么是现实?这些问题的答案尽在本章中。

第八章 智能之未来(211)

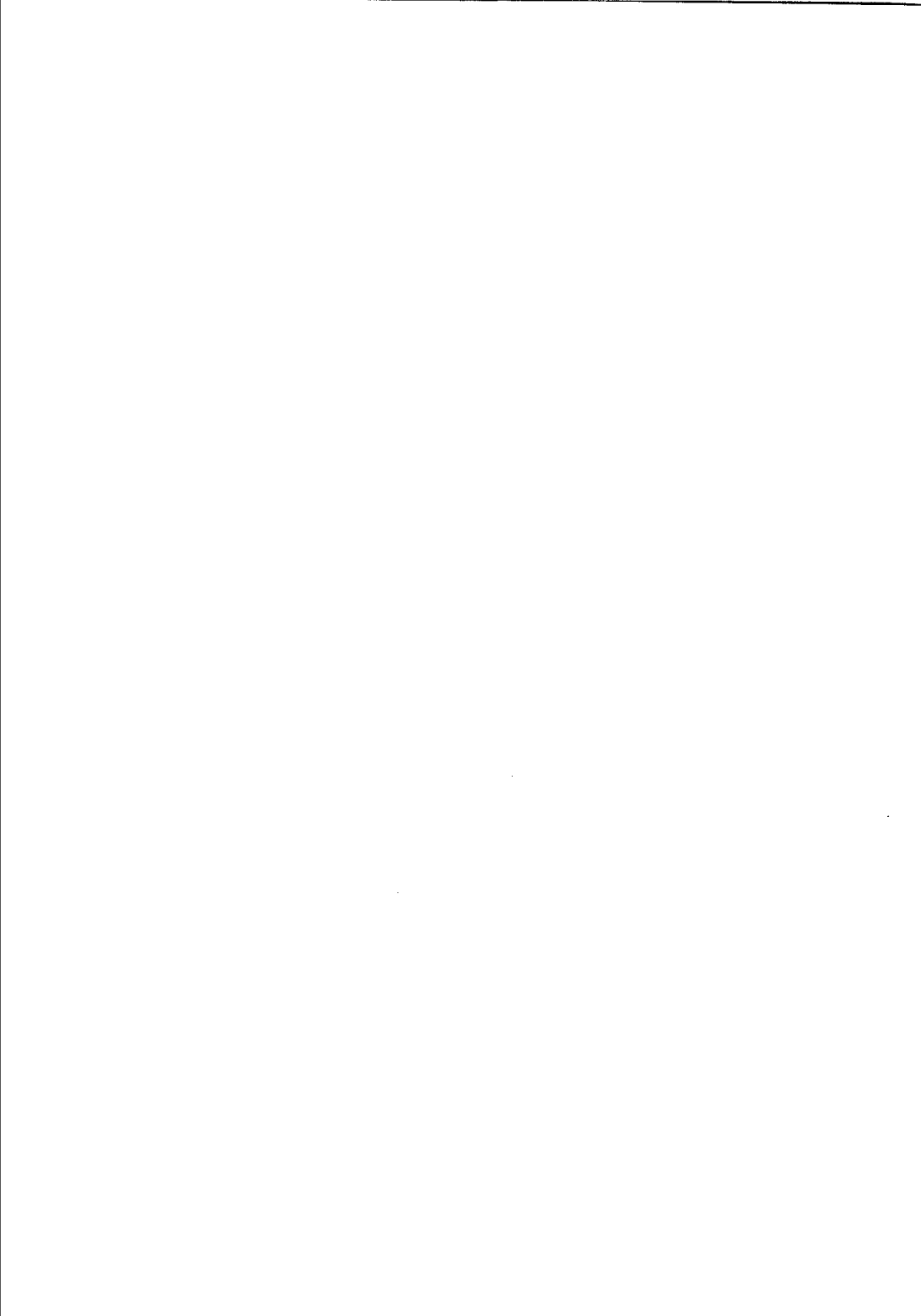
很难预言一项新技术的最终用途，但设想一下未来智能机器的样子还是有所裨益的。人们能研制出智能机器吗？如果可以，那它们将会是什么样子呢？它们将会被用来干什么呢？这项技术是不是很危险，会不会危害到人类并对个人自由构成威胁？智能机器最显而易见的应用范围将是什么？智能机器对我们生活的最终影响将会是什么呢？

结语(243)**附录 可检验的预言**(247)

第一章

人工智能

Artificial Intelligence



当我 1979 年 6 月从康奈尔大学毕业时，我对自己的生活没有任何长远的打算。获得电机工程学位后，我在位于俄勒冈州波特兰市新建的英特尔工业园找到了一份工程师的工作。当时，微型计算机工业方兴未艾，而英特尔是这一领域的核心。单片机是我们当时的主导产品（由于英特尔发明了微处理器，因此在那时，把整台计算机建立在单个电路板上的想法就已经成为了可能）。而我的工作就是分析并修复由其他工程师发现的单片机上的错误。那期间，我创办了一份业务通讯，并经常外出旅行，有机会见识各种客户。尽管很思念在辛辛那提工作的大学女友，但那时的我年轻，过得很快活。

几个月后，我遇到了一件改变我人生目标的大事。那是一本当年 9 月新出版的《科学美国人》(Scientific American) 杂志，是有关大脑研究的专刊。它重又唤起了我自小就开始的对大脑的兴趣。这本杂志令人着迷。里面的文章涉及大脑的组织、发展和化学特征，专事视觉、运动及其他此类组织的神经机制，以及精神错乱的生理基础等内容。我认为它是《科学美国人》里最棒的一期。一些与我交往过的神经学家也表示，这期杂志在帮助他们做出职业选择的时候起到了重要作用。这一点和我一样。

这一期的最后一篇文章题为《有关大脑的思考》，作者是弗朗西斯·克瑞克(Francis Crick)，他曾是 DNA 结构的发现人之一，并且他那时就已经将其天才思考投入到了对大脑的研究之中。克瑞克认为，尽管一直以来积累了大量有关大脑的详尽知识，但大脑的工作原理对于人们来说仍是一个难解的谜。一般来说，科学家不会写他们不知道的事情，但克瑞克就像那个指出皇帝没有穿衣服的小男孩一样，对此并不在意。他认为神