

# 暗知识：机器认知如何颠覆商业和社会

作者：[美]王维嘉

## 版权信息

COPYRIGHT

书名：暗知识：机器认知如何颠覆商业和社会

作者：【美】王维嘉

出版社：中信出版集团

出版时间：2019年4月

ISBN：9787508692982

本书由中信出版集团授权得到APP电子版制作与发行

版权所有·侵权必究

# 导读

一直以来人类的知识可以分为两类：“明知识”和“默知识”（Tacit Knowledge，又称默会知识）。明知识就是那些可以用文字或公式清晰描述和表达出来的知识。默知识则是个人在感觉上能把握但无法清晰描述和表达的知识，也即我们常说的“只可意会，不可言传”的那类知识。人类发明文字以来，积累的知识主要是明知识，因为只有明知识才可以记录和传播。直到大约70年前，人类才意识到默知识的存在。今天，人工智能，特别是其中的一个重要流派——神经网络，突然发现了海量的、人类既无法感受又无法描述和表达的“暗知识”——隐藏在海量数据中的相关性，或者万事万物间的隐蔽关系。这些暗知识可以让我们突然掌握不可思议的“魔力”，能够做很多过去无法想象的事情。本书就是要清楚阐述机器学习发掘出了什么样的暗知识，为什么机器能够发现这些暗知识，以及这些暗知识对我们每个人会有什么影响。

本书分为三个部分。

第一部分包括第一、二、三章，其中第一章里我们发现AlphaGo（阿尔法围棋）给我们带来的最大震撼是人类完全无法理解机器关于下棋的知识。这个发现迫使我们重新审视人类对于“知识”的所有观念。这一章回顾了2500年来人类所熟悉的明知识和直至大约70年前才注意到的默知识。近几十年的脑神经科学的研究成果让我们对知识的本质有了更清楚的认识，也回答了为什么人类既无法感受，也无法理解机器发现的那些暗知识。这一章还分析了明知识、默知识和暗知识之间的区别，讨论了为什么暗知识的总量将远远超过人类能掌握的所有知识。

第二章介绍了机器是怎样学习的，能学习哪些知识，同时介绍了机器学习的五大流派以及各流派从数据中挖掘知识的方法。

第三章则重点介绍了目前机器学习中最火的神经网络，包括神经网络的基本工作原理和目前在商业上应用最广的几种形态，以及各自适用的领域。有了这些基础就可以判断AI（人工智能）在各个行业的商业机会和风险。也只有理解了这些原理，才能真正理解暗知识的特点。为易于阅读和照顾不同读者的需求，在这一章中我们尽量用通俗的语言解释这些工作原理，而把精确的技术原理介绍放在附录里。

第二部分（第四、五章）讨论了AI对商业的影响。我们将看到机器发掘出来的暗知识对我们生活的直接影响。对于想把握AI商业趋势的读者来说，这部分的内容至关重要。其中，第四章描述了当前的AI产业生态，第五章详尽探讨了哪些行业将面临AI的颠覆，以及在不同行业的投资机会和陷阱。

第三部分（第六、七章）的内容是AI对未来的影响。第六章重点讨论目前还没有商业化的，但可能更深刻影响我们的一些神奇的AI应用。第七章讨论了机器人和人的关系：机器能在多大程度上取代人的工作，会造成哪些社会问题（例如大面积失业）。这两章的主要目的是开脑洞，探讨那些我们今天可能还看不到的更深远的影响。本章也试图回答人类的终极恐惧：机器人最终会控制人类吗？

本书的各个章节前后连贯，但也可以跳着读，对于那些只对商业感兴趣的读者，可以跳过第二、三章直接读第四、五章。

笔者在美国斯坦福大学读博士期间做过人工智能研究，后来在硅谷和中国创办高科技公司，目前在硅谷专注于投资人工智能。每年访问调研上千家硅谷和中国的科技公司，接触顶级大学最前沿的研究，这些都有助于笔者从大量的实践中提炼出自己对行业的原创的分析和洞见，而不是人云亦云。

笔者长期对人类如何获得知识感兴趣，在投资、研究和写作AI的过程中，发现了暗知识这样一个人类以往未曾发现的领域。这个概念的提出一定会引起争议，笔者欢迎读者的批评并期待在批评和讨论中进一步深化在这方面的认识。

本书的目标读者是企业 and 政府工作人员及其他知识阶层，包括学生。暗知识对人类的影响刚刚开始。从暗知识这个新视角出发，可以更深刻地理解这次AI巨浪。这波巨浪可能超过互联网，许多行业都会深受影响。本书希望能回答“AI对我的行业和职业会有什么影响”。只有把AI的技术、趋势和应用深入浅出地讲清楚，读者才可能举一反三，理解AI对自己的影响。本书从笔者自己的投资实践出发，希望能为在AI时代进行投资提供一些参考。在AI飓风里泥沙俱下，鱼龙混杂，会有大量的炒作，读完本书可以帮助读者辨别真伪，不会被轻易忽悠。在今后5~10年，不论是风险投资/私募股权投资还是在公开股票市场投资都需要有这样的辨别能力。本书最后在讨论人工智能对整个社会的影响时也提出了一些未经检验的建议。

每当读到市面上科技类的书籍时，常被那些含混不清的描述所困扰。当年在斯坦福大学上课时留下的最深印象就是那些学科的开山鼻祖对自己学科理解之深入。他们能用最简单的方式把最深奥的道理讲明白，让听课的学生一下子就能理解一门学科的核心概念，而且一辈子不会忘记。从那以后，笔者就坚信，如果学生没听懂，一定是老师没讲明白。这本书希望用最通俗易懂的语言介绍暗知识和AI。任何具有高中以上学历的读者如果有没读懂的地方，一定是因为笔者没有写明白。

今天每个人都面对海量的信息和知识，如何让读者花最少的时间获取大量的信息和知识成为一个挑战。笔者最欣赏的文章和书籍是那些没有一句多余的话的，这也是笔者写作本书的目标之一。本书希望能够做到读者在机场书店买了这本书后能在下飞机前读完，而且读完之后可以清晰地判断这场技术大浪对自己的影响。

王维嘉

2019年1月13日于硅谷

# 序言

## “暗知识”和现代社会

自2017年AlphaGo大胜柯洁，人工智能即将碾轧人类的话题便进入大众视野，迅即引起普遍的狂热和焦虑。我认为，王维嘉这本《暗知识：机器认知如何颠覆商业和社会》的出版，是对这种情绪的有效清醒剂和解毒药。

说这本书是清醒剂，是因为它极为简明清晰地叙述了人工智能的科学原理及其技术实现，无论是神经网络结构，其自我学习的过程，还是深度学习和卷积机制，《暗知识：机器认知如何颠覆商业和社会》比现在出版的任何一本书都讲得更清楚、易读。以人类认知为背景来解读人工智能，正好可以为当前人工智能领域中泛起的非理性狂热降温。其实，早在20世纪60年代，控制论创始人维纳（Norbert Wiener）的学生阿比布（Michael Arbib）在《大脑、机器和数学》一书中，已经清晰地叙述了神经网络数学模型和学习机原理，并讲过这些原理有助于我们“从‘机器’中‘赶走鬼魂’”。阿比布讲的“机器”是指大脑的记忆、计算和学习等功能，它们自笛卡儿以来被视为机器的有机体（生物），“鬼魂”则是指生物的本能和学习能力。而王维嘉的《暗知识：机器认知如何颠覆商业和社会》一书，“赶走”的不是以往所说的有机体的神秘性，而是对人工智能研究和可能性的想象中的“鬼魂”，即误以为当神经元的连接数量接近于人脑时，它们会涌现出如人类那样的自我意识和主体性等。

人工智能的神经网络系统能做什么？如上所说，早在它被做出来以前，数学家已经证明，无论神经网络多么复杂，它等价于有限自动机；而一个能和环境确定性互动（自耦合、反馈和自我学习）的有限自动机（神经网络），只不过是某一种类型的图灵机（通用计算机）。

也就是说，人工智能革命之基础——神经元的自我学习及其与环境互动所能达到的极限，都超过图灵机的行为组合。从20世纪下半叶至今，伴随着人工智能的快速、高度发展，关于它能否在未来某一天具有意识的讨论，一直是在电脑和人脑差别的框架中展开的。我认为，只要发展出相应的数学理论，就能了解神经网络学习已做出的和可能做的一切。但有一点是毫无疑问的，它不可能具有自我意识、主体性和自主性。

为什么说这本书是解毒药？因为维嘉在解释为什么人工智能可以比人更多、更快地掌握知识（能力）时，把人工智能所掌握的信息定义为“暗知识”，从而可以得出清晰的理论表述。我们首先要弄明白什么是知识，知识就是人获得的信息。而人利用信息（知识）离不开获得信息和表达信息两个基本环节，人获得信息是用感官感知（即经验的），表达信息是通过符号（语言）和对符号结构之研究（符号可以是非经验的）。这样，他根据“可否感知”和“可否表达”，把人可利用的知识分为如下四种基本类型：

第一，可感知亦可表达的知识。它包括迄今为止所有的科学和人文知识。

第二，不可感知但可表达的知识。任何经验的东西都是可感知的，不可感知的就是非经验的。有这样的知识吗？当然有。以数学为例，抽象代数的定理是正确的知识，但可以和经验无关。人之所以为人，就在于可以拥有纯符号的知识，它是理性的重要基础。

第三，可感知但不可表达的知识。它包括人的非陈述性记忆和“默会知识”。

第四，不可感知亦不可表达的知识。这就是当前神经网络通过学习掌握的知识。维嘉将这类大大超出了个人所能记忆和学习的知识称为“暗知识”。“暗知识”的提出，不仅是一项哲学贡献，也为当前盛行的科学乌托邦提供了一剂解毒药。

20世纪社会人文研究最重要的成就，就是发现“默会知识”和市场的关系。人类可共享的知识都是可以用符号表达的知识，但它不可能包含每个人都具有的“默会知识”。经济学家利用“默会知识”的存在，证明了基于理性和科学知识的计划经济不可能代替市场机制。一个充分利用人类知识的社会，一定是立足于个人自主、互相交换自己的能力和知识所形成的契约组织。忽视所有个人具有的“默会知识”，把基于理性和可表达的知识设计出的社会制度付诸实践，会出现与原来意图相反的后果。哈耶克称这种对可表达知识的迷信为“理性的自负”。今天随着大数据的应用，这种理性的自负再一次出现在人工智能领域。而“暗知识”的提出，扩大了不能用符号表达知识的范围，进一步证明了哈耶克的正确性。所以，我说这本书是对当前理性自负的有效解毒药。

维嘉在书中提出的另一个有意义的问题是“暗知识”会在何种程度上改变现代社会。正如该书副标题所说，这种新型知识大规模的运用，将会导致大量拥有专门知识和技能的失业、一批又一批的行业消失，甚至连医生专家都可能被取代。姑且不论这种预测是否准确，有一点是肯定的，即人工智能必定会极大地改变我们赖以生存的社会。那么，它会把人类社会带到哪里去？这正是人工智能革命带来的普遍焦虑之一。人工智能对城市管理和对每个人隐私的掌握，是否会导致个人自由和隐私的丧失？由大数据和人工智能高科技管理的社会，还是契约社会吗？

现代社会和传统社会的本质不同就在于其高度强调个人的主体性和创造性，任何信息的获得、表达和应用都离不开个人的主体性和创造性。我认为，人工智能可以具有掌握“暗知识”的能力，但它不可能具有自我意识，当然亦无所谓主体性，它只能被人所拥有。因此，一个能允许知识和技术无限制进步的社会，仍然是建立在个人契约之上的。也就是说，无论科学技术发展到什么程度，现代社会的性质不会因而而改变。

然而，我认为，人工智能或许会使现代社会的科层组织的形式发生改变。为什么现代社会需要科层组织？众所周知，现代社会除了由法律和契约提供组织框架以外，还必须向所有人提供不同类型的公共事务服务，如治安、交通设施、教育、医疗等。为此就要设立处理不同类型事务的专门机构来管理社会，如军队和政府科层组织。科层组织之间功能的实现和协调，要利用符号表达的共享知识，因此，随着现代社会的复杂化，必定出现技术官僚的膨胀。而人工智能革命和“暗知识”的运用，必定会向社会管理层面深入。如果它运用得不好，会使现代社会生长出超级而无能的官僚机构的毒瘤；如果它运用得好，可以促使人更好地发挥自主性和创造性，甚至可以取代科层管理中不必要的机构。因此，我认为人工智能将会在这一层面给现代社会带来巨大影响。

科层组织的形成和理性化的关系，是韦伯分析现代社会的最重要贡献。在未来，随着人工智能对“暗知识”的掌握和运用向社会管理渗透，甚至替代，将会证明韦伯这一重要论断不再成立。可惜的是，维嘉的《暗知识：机器认知如何颠覆商业和社会》一书忽略了人工智能革命和现代社会官僚化关系的讨论。科层组织的设立是基于理性（共享知识），人工智能擅长的是掌握“暗知识”，如果从事社会公共事务管理的人员可以被掌握“暗知识”的人工智能取代，科层组织还有存在的必要吗？或者，它将以什么样的新形式存在？如果不再需要科层组织，未来无政

府的现代社会将如何运行？这正是我们应该关注的，它需要人文和科学两个领域的对话。

金观涛

2019年2月

## 寄语

□

译：我非常高兴推荐这本书。这本书对机器学习的发明带来的下一场工业革命进行了详尽的分析。我希望这个技术将被用来使人类的生活更美好、更和平，并不再有战争。——伯纳德·威德罗，2019年2月26日于斯坦福

# 第一章

## 横空出世——暗知识的发现

### 导读

正当人类自以为掌握了关于这个世界的海量知识时，一种能够自我学习的机器给了我们当头一棒：机器发现了一类人类既无法感受，也不能理解的知识。这类知识的发现，逼迫我们重新审视过去所有关于知识的观念。我们回顾了2500年来在这个问题上的争论：知识是通过经验得到的还是通过推理得到的？直到大约70年前人们才注意到那些“只可意会，不可言传”的默知识的重要性。但这些争论在最新的脑科学研究结果面前都显得肤浅和苍白。最近几十年的科学研究确认了认知的基础是大脑神经元之间的连接。有了这个基础，我们就很容易理解为什么有些知识无法表达，也才能明白为什么人类无法理解机器刚刚发现的这些暗知识。在此基础上，我们终于可以清晰地区分这样三类知识：人类能掌握的明知识和默知识以及只有机器才能掌握的暗知识。

## 骄傲的人类

也许是由于几十万年人类远古祖先某个基因的突变，人们开始可以把一些有固定意思的发音片段组装成一个能表达更复杂意思的发音序列。这些发音片段今天我们叫作“单词”，这个表达特定内容的发音序列今天我们叫作“句子”。这种“组装”能力使人类用有限的单词可以表达几乎无穷多种意思，语言诞生了。有了语言的复杂表达能力，人类的协作能力开始迅速提高，可以几十人一起围猎大型动物，很快人类就上升到地球生物链的顶端。作为记录语言的符号——文字的发明可以让人类更方便地传播、记录和积累经验。任何一个地方的人类偶然发现的关于生存的知识都会慢慢传播开来。一万年以前，农业起源于今天的埃及、叙利亚和伊拉克的肥沃新月带，这些种植经验在几千年中传遍全世界，随之而来的是人类迅速在地球所有适宜农耕的角落定居繁衍。

随着定居的人类数量的增加，人类的组织开始变得更大更复杂，从亲缘家族到部落，到城邦，再到国家。大规模的复杂组织可以开展大规模的复杂工程，如建设城市、庙宇和大规模灌溉系统。这些大规模工程需要更多的天文和数学知识。世界上几乎所有的古老文明都积累了许多天文知识，但只在希腊半岛诞生了现代科学的奠基石——数学。欧几里得（Euclid，公元前330—前275）在公元前300年总结了他前面100年中希腊先哲的数学成果，写出了人类历史上最伟大的书之一《几何原本》（Elements）。这本书在中世纪由波斯裔的伊斯兰学者翻译成阿拉伯文，又从阿拉伯传回文艺复兴前的欧洲，直接影响了从哥白尼（Nicolaus Copernicus，1473—1543）到牛顿（Issac Newton，1643—1727）的科学革命。

发轫于16世纪的科学革命的本质是什么？是发现更多的知识吗？是创造出更多的工具吗？都不是。科学革命的本质是找到了一个可靠的验证知识的方法。

最能体现科学革命本质的就是天文学家开普勒（Johannes Kepler，1571—1630）发现三定律的过程。最初，在作为主流的托勒密（Ptolemy，90—168）地心说越来越无法解释天体观测数据时，哥白尼提出了日心说，用新的模型解释了大部分过去无法解释的数据。与伽利略（Galileo Galilei，1564—1642）同时代的天文学家第谷·布拉赫（Tycho Brahe，1546—1601）没有接受哥白尼的日心说，他提出了“月亮和行星绕着太阳转，太阳带着它们绕地球转”的“日心—地不动”说。遗憾的是，他倾尽毕生心血观察了20年的天文数据，直到去世都始终无法让观测到的数据与自己的模型相吻合。

在第谷去世后，第谷的助手开普勒拿到了他的全部数据，开普勒完全接受了哥白尼的日心说。他为了让数据与日心说完全吻合，把哥白尼的地球公转的圆形轨道修正为椭圆轨道，太阳在椭圆的一个焦点上。这就是开普勒第一定律。他用相同的方法发现了其他两个定律。开普勒三定律不仅完美解释了第谷的所有观测数据，并且能够解释任何新观测到的数据。

这个发现过程有三个步骤：第一，积累足够的观测数据（第谷20年的观测数据）；第二，提出一个先验的世界模型（哥白尼的“日心说”）；第三，调整模型的参数直至能够完美拟合已有的数据及新增数据（把圆周轨道调整为椭圆轨道，再调整椭圆轴距以拟合数据）。验证了这个模型有什么用？最大的用处就是可以解释新的数据或做出预测。在这里开普勒三定律就是新发现的知识。发现知识的可靠方法就是不断修改模型使模型与观测数据完全吻合。

上面这三个步骤奠定了现代科学的基本原则，正式吹响了科学革命的号角，直接导致了后来的牛顿万有引力的发现，一直影响到今天。

过去500年中人类对世界的认识突飞猛进，今天大到宇宙，小到夸克都似乎尽在人类的掌握之中。人类可以上天、入地、下海，似乎无所不能。人类有了“千里眼”“顺风耳”，甚至开始像“上帝”一样设计新的物种，并企图改变人类进化的进程。人类有理由相信没有什么知识是不能理解的，也没有什么知识是不能被发现的……直到2016年3月15日。

## 天才的哽咽

2016年3月15日，美国谷歌公司的围棋对弈程序AlphaGo以五局四胜的成绩战胜世界围棋冠军韩国选手李世石。一时间这个消息轰动世界，全世界有28亿人在关注这场比赛，在中国更是引起极大的轰动。人们感觉AlphaGo就像从石头缝里蹦出来的孙悟空一样，完全无法理解一台机器如何能够打败世界围棋冠军。围棋历来被认为是人类最复杂的游戏之一。围棋每一步的可能的走法大约有250种，下完一盘棋平均要走150步，这样可能的走法有 $250^{150}=10^{360}$ 种，而宇宙从诞生到现在才1017秒，即使是现在世界上最快的超级计算机，要想把所有走法走一遍，计算时间也要比宇宙年龄都长。即使排除了大部分不可能的走法也是大到无法计算。机器是怎样学会这么复杂的棋艺的？

这场比赛后，世界排名第一的棋手柯洁在网上说：“AlphaGo胜得了李世石，胜不了我。”而2017年5月28日，棋手柯洁以0：3完败AlphaGo，彻底击碎了人类在这种复杂游戏中的尊严。赛后，这位天才少年一度哽咽，在接受采访时柯洁感叹，AlphaGo太完美，看不到任何胜利的希望。他流着眼泪说：“我们人类下了2000年围棋，连门都没入。”中国棋圣聂卫平更是把AlphaGo尊称为“阿老师”，他说：“AlphaGo的着数让我看得如醉如痴，围棋是何等的深奥和神秘。AlphaGo走的顺序、时机掌握得非常好。它这个水平完全超越了人类，跟它挑战下棋，只能是找死。我们应该让阿老师来教我们下棋。”他还说：“阿老师至少是20段，简直是围棋上帝。”

当人们以为这是对弈类程序的高峰时，AlphaGo的研发团队DeepMind（谷歌收购的人工智能企业，位于伦敦）团队再度碾轧了人类的认知。2017年12月，DeepMind团队发布了AlphaGo Zero（阿尔法围棋零）。AlphaGo Zero使用了一种叫作“强化学习”的机器学习技术，它只使用了围棋的基本规则，没有使用人类的任何棋谱经验，从零开始通过自我对弈，不断地迭代升级，仅仅自我对弈3天后，AlphaGo Zero就以100：0战胜了此前击败世界冠军李世石的AlphaGo Lee版本。自我对弈40天后，AlphaGo Zero变得更为强大，超过了此前击败当今围棋第一人柯洁的AlphaGo Master（大师版），这台机器和训练程序可以横扫其他棋类。经过4个小时的训练，打败了最强国际象棋AI Stockfish，2个小时打败了最强将棋（又称为日本象棋）AI Elmo。

AlphaGo Zero证明了即使在最具有挑战性的某些领域，没有人类以往的经验或指导，不提供基本规则以外的任何领域的知识，仅使用强化学习，仅花费很少的训练时间机器就能够远远超越人类的水平。

## 机器发现了人类无法理解的知识

AlphaGo Zero给我们的震撼在于人类2000多年来一代代人积累的一项技艺在机器眼里瞬间变得一文不值！为什么会这样？围棋中的可能走法比宇宙中的原子数都多，而人类2000多年中高水平对弈非常有限，留下记录的只有几万盘。这个数字和所有可能走法比，就像太平洋里的一个水分子。而AlphaGo Zero以强大的计算能力，在很短的时间里探索了大量的人类未曾探索过的走法。人类下棋的路径依赖性很强，人生有

限，想成为高手最稳妥的办法是研究前人的残局，而不是自己瞎摸索。但AlphaGoZero在下棋时，不仅一开始的决策是随机的，即使到了大师级后，也故意随机挑选一些决策，跳出当前思路去探索更好的走法，新发现的许多制胜走法都是人类从未探索过的，这就是很多走法让聂卫平大呼“看不懂”的原因。

AlphaGo Zero给我们的震撼在于三个方面：首先，人类能发现的知识和机器能发现的相比，就像几个小脚老太太走过的山路和几百万辆越野车的山路。越野车的速度就是计算机和AI芯片处理速度，目前继续以指数速度在提高。其次，和机器可能发现的知识相比，人类知识太简单、太幼稚，机器谈笑风生，比人不知道高到哪里去了。最后，机器发现的知识不仅完全超出了人类的经验，也超出了人类的理性，成为人类完全无法理解的知识。

2500年前最有智慧的希腊哲人苏格拉底（Socrates，公元前469—前399）终其一生得出一个结论：“我唯一知道的是我什么都不知道。”他的学生柏拉图（Plato，公元前427—前347）认为我们感官观察到的世界只是真正世界的影子而已。18世纪伟大的哲学家康德也仰望星空，发出了“我们到底能知道什么”的千古之问。但古代哲人只能模糊地感觉到人类认识的局限。今天，AlphaGo Zero不仅清晰、具体地把他们的疑虑变成了铁的事实，而且先哲怎么也想不到人类的认识能力是如此有限！

你会质疑说：这不算什么震撼吧，人类早就知道我们已知的很少，未知的很多。但这个下围棋的例子告诉你：已知的是几万盘残局，未知的是10360种可能走法，两者相差几百个数量级！（不是几百倍，是几百个数量级，一个数量级是10倍。）

你学过概率和统计，继续不服：我们早就知道组合爆炸。没错，但我们知道未知的组合爆炸里有比人类已经获得的知识高深得多的知识吗？AlphaGo Zero是第一次活生生地证明了这点。听说过火山爆发和在现场看到的感受能一样吗？

当然最震撼的就是第三个方面。我们也许知道我们不知道很多，甚至能用逻辑推断出未知知识里有比已知知识更高深的知识，但我们怎么也想不到这些知识是人类根本无法理解的。这是人类历史上第一次遇到这样的问题，我们给自己造了个“上帝”！这件事对哲学和认识论的冲击空前，人类突然不知所措，影响还在发酵，后果不可估量。

“理解”的意思是要么能用感觉把握事物间的关系，要么能用概念把经验表达出来，或者用逻辑把事物间的关系表达出来。无法理解就等于既无法感受又无法表达。

也就是说，机器发现了人类既无法感受也无法表达的知识。用更通俗的话说就是，机器发现了那些既无法“意会”又无法“言传”的知识。

一个无法理解的知识的表现形式是什么样的？如果无法理解又怎么判断它就是知识？当我们想回答上面的问题时，我们发现必须重新审视什么是“知识”。人类过去几千年是怎样获得知识的，获得了什么样的知识？就像每次科学上的重大发现都要迫使我们重新审视过去习以为常的观念一样，今天机器的震撼让我们必须重新审视过去所有关于“知识”的基本理念。

人类获得知识的行为就是认知。过去我们对世界的认识局限主要来自观察能力。在望远镜发现之前，第谷根本无法观测行星运动，当然更谈不上记录数据，也不会有后来的开普勒定律和牛顿万有引力定律。在显微镜发明之前，我们不可能发现微生物，一切关于细胞和基因的发现都无从谈起。今天谁能花1000万美元买一台冷冻电镜，谁就可以看到别人看不到的分子晶体结构，就可以经常在《自然》（Nature）杂志上发表文章。随着新的观察仪器的出现和已有观察仪器的改进，我们对世界的认识还会不断深入。

我们对世界认识的第二个局限来自解释能力。所谓解释能力就是发现事物间的因果关系或者相关性并能够表达出来。即使我们能观察到许多现象，如果我们无法解释这些现象则还是无法从这些观察中获得知识。例如第谷虽然有大量观测数据，但终其一生没有找到能解释数据的正确模型。又如我们观察到人有语言能力而黑猩猩没有，但不知道为什么，仅仅是知道这个现象而已。

人类几千年来关于知识的争论正是围绕着“观察”还是“解释”展开的。

## 理性主义和经验主义之争

自从5000年前两河流域的苏美尔人发明了人类最早的文字——楔形文字以来，人类一直在记录和积累知识。但直到2500年前希腊人才开始系统地研究关于知识的学问。在这个问题上，一直有两大流派：理性主义和经验主义。

第一个开启了理性主义的人是苏格拉底。人类此前的大部分“知识”要么从宗教教义中来，要么从传统习俗中来。人们从生下来就不加怀疑地接受了这些东西。而苏格拉底则要一一审视这些东西。苏格拉底说我们都希望有一个“好”的人生，但到底什么是“好”什么是“坏”呢？不去质疑，不去探究你怎么知道呢？所以探究和道德是不可分割的，不去探究我们身边的世界不仅是无知而且是不道德的，所以他的结论是：一个未经探究的人生根本就不值得过。他平时没事就跑到大街上拉住人诘问：“什么是正义？”“什么是善？”“什么是美？”每当人们给他个定义时，他都能举出一个反例。他这种探究思辨影响了无数代人。后来当他的学生柏拉图把“人”定义为“没有毛的双足动物”时，当时的另一位哲学家提奥奇尼斯马上拿来一只拔光了毛的鸡说：“大家请看柏拉图的‘人’！”经过一生的探究，苏格拉底得出结论“我唯一知道的是我什么也不知道”。苏格拉底式思辨震撼了当时的社会，传统势力认为这样会搞乱人心，当政者用“腐蚀青年思想罪”判处他死刑，他最终饮毒酒身亡。他一生全部用来和人辩论，没有留下任何著作。幸亏他的学生柏拉图把老师的辩论编辑成了传世之作《对话录》。正是苏格拉底开启了通过逻辑思辨来验证知识的希腊传统。

如果说是苏格拉底开了理性主义的先河，他的弟子柏拉图就是理性主义集大成的鼻祖。苏格拉底的思辨主要集中在道德哲学领域，探究什么是“公平”和“善”。而柏拉图则对他的先辈毕达哥拉斯（Pythagoras，约公元前570—前495）开创的数学传统深为折服。柏拉图的学说深受数学严格推理的影响。他甚至在他创办的学园门口挂了个牌子：“不懂几何者不得入内。”柏拉图学说的核心是“理想原型”。他说，世界上每一条狗都不一样，我们为什么认为它们都是狗？人类心中一定早有一个关于狗的理想原型。我们知道三角形的内角之和等于180度，但我们从未见过一个完美的三角形。他认为人类的感官无法触及这些理想原型，我们能感受到的只是这些理想原型的失真拷贝。真实世界就像洞穴外的一匹马，人类就像一群背对着洞口的洞穴人，只能看到这匹马在洞穴壁上的投影。柏拉图奠定了理性主义的两大基础——知识（理想原型）是天生的；感官是不可靠的，并由此推出理性主义的结论：推理而不是观察，才是获取知识的正确方法。

亚里士多德（Aristotle，公元前384—前322）17岁进入柏拉图的学园当学生，当时柏拉图已经60岁了。亚里士多德在学园里待了20年，直到他的老师柏拉图去世。亚里士多德对老师非常尊敬，但他完全不同意老师的“理想原型”是先天的。他认为每一条狗都带有狗的属性，观察了许多狗之后就会归纳出狗的所有属性。这个“理想原型”完全可以通过后天观察获得，而不需要什么先天的假设。柏拉图酷爱数学，而亚里士多

德喜欢到自然中去观察植物和动物。两人的喜好和经历是他们产生分歧的重要原因之一。亚里士多德认为：知识是后天获得的，只有通过感官才能获得知识。正是亚里士多德开了经验主义的先河。

经验主义这一派后世的著名代表人物有英国的洛克（John Locke, 1632—1704），贝克莱（George Berkeley, 1685—1753）和休谟（David Hume, 1711—1776），贝克莱认为人生下来是一张白纸，所有的知识都是通过感官从经验中学来的。但理性主义则认为，经验根本不可靠。英国哲学家罗素（Bertrand Russell, 1872—1970）有个著名的“火鸡经验论”。火鸡从生下来每天都看到主人哼着小曲来喂食，于是就根据经验归纳出一个结论：以后每天主人都会这样。这个结论每天都被验证，火鸡对自己的归纳总结越来越自信，直到感恩节的前一天晚上被主人宰杀。理性主义者还问：眼见为实吗？你看看图1.1中的横线是水平的还是倾斜的？

图1.1视错觉图（图中所有横线都是水平的）

理性主义的后世代表人物则有法国的笛卡儿（Rene Descartes, 1596—1650）和德国的莱布尼茨（Gottfried Leibniz, 1646—1716）。笛卡儿有句名言“我思，故我在”，我的存在这件事不需要经验，不需要别人教我，我天生知道。莱布尼茨是和牛顿一样的天才，他和牛顿同时发明了微积分，也是二进制的发明人，还发明了世界上第一台手摇计算器。他认为世界上每个事物都包含了定义这个事物的所有特性，其中也包含了和其他事物的关系。从理论上我们可以用推理的方法预测全宇宙任何一点，过去和未来任何时间的状态。

理性主义认为，感官根本不靠谱，最可靠的是理性，基于公理严格推导出来的几何定理永远都不会错。理性主义找出更多的例子来说明人类的最基本的概念是天生的。例如自然数，我们怎么学会“1”这个概念的？拿了一个苹果告诉你“这是一个苹果”；又给你拿了个橘子告诉你“这是一个橘子”。但苹果是苹果，橘子是橘子，两者没关系，你怎么就能抽象出“1”这个概念来呢？又比如我们可以根据直角三角形的特点推导出勾股定理，又进一步发现世界上居然有无法用分数表达的无理数。这种革命性的发现完全不依赖感觉和经验。小孩一出生就知道这个球不是那个球，这条狗不是那条狗，这个“同一性”是理解世界最基本的概念，没人教他。

我们注意到理性主义有一个隐含的假设，就是因果关系。在莱布尼茨的世界里，一件事会导致另外一件事，所以才有可能推导。经验主义当然不服，休谟就问，一件事发生在另外一件事之后，两者未必有因果关系。譬如我把两个闹钟一个设在6:00，一个设在6:01，能说后面的铃声响了是前一个造成的吗？理性主义不仅认为事物间有因果关系，而且认为通过逻辑推理可以得到很多知识。譬如归纳推理：太阳每天早上都会升起。但休谟就质问：你能像证明数学定理一样证明太阳明天会升起吗？不能吧。那能观察吗？明天还没到来显然不能观察，那你凭什么说明天太阳一定升起，我要说明天不一定升起错在哪里了？我们看到休谟挑战的是归纳背后的假设：事物运动规律不变，在这里就是说地球和太阳系的运动不会改变。休谟最后说，物理世界没什么因果，没什么必然，你最多能根据以往的经验告诉我：明天早上太阳还可能升起。

这两派从17世纪吵到18世纪，这时候在德国偏僻的海德堡出现了一个小个子乡村秀才。他说，你们双方似乎都有道理，我来整合一下看看。他就是哲学史上最影响力的康德（Immanuel Kant, 1724—1804）。康德说，没错，我们当然要通过感官去理解世界。但我们对事物的理解包括这个事物的具体形态和它的抽象概念。譬如眼前这本书，一本书的具体形态千变万化，但“书”这个概念就是指很多页有字的纸装订在一起的一个东西。我们说“面前有这本书”的意思到底是什么？那至少要说现在几月几日几点几分，在某市某区某小区几号楼几号房间的哪个桌子上有这本书，也就是理解一个具体的东西离不开时间和空间的概念。但谁教给你时间和空间了？你妈从小教过你吗？你教过你孩子吗？好像都没有，我们好像天生就懂。所以康德说，你看，必须有这些先天就有的概念你才能理解世界。我们好像天然知道“书”是个“东西”，“东西”是一种不依赖我们的独立存在。谁教给我们“东西”这个概念的？没人，好像又是天生就懂吗？康德整合了经验主义和理性主义，他的一句名言是“没有内容的思维是空洞的，没有概念的感知是盲目的。只有把两者结合我们才能认识世界”。

在2500年的辩论中，经验主义当然不会否认数学中通过严格推理得出来的结论的可靠性，理性主义也不得不承认认知物理世界离不开感官。那么这场打了2500年的嘴仗到底在争什么呢？问题出在理性主义者企图把数学世界里证明定理这样的绝对可靠性推广到物理世界，也即他们企图找到一个检验知识的普遍的标准，能够适用于所有领域。数学（例如几何学）是建构在公理之上的一个自洽而完备的系统（至少对自然数和几何是如此）。所谓自洽就是说，在这个系统里只要从公理出发就不会推导出互相矛盾的结论；所谓完备就是说，在这个系统里任何一个命题都是可以证实或证伪的。而亚里士多德时代的自然科学的可靠性判断标准是“观察与模型符合”，即观察到的自然现象和事先假设的模型的预测结果相符合。这种物理真实性的判断标准和数学中的判断标准完全不同。所以经验主义觉得硬要把数学中的可靠性标准搬到自然科学中来不适用，或者说经验主义认为在自然科学领域只能依赖感官。因此这场争论是不对称的：理性主义要从数学攻入自然科学，而经验主义死守自然科学的阵地。双方撕扯不清的另一个原因是谁都不知道感官和认知的本质是什么，或者说知识的本质是什么。双方根据自己的猜测和假设激烈辩论，一直到20世纪50年代人们对大脑的研究才取得突破。

## 知识的生物学基础——神经元连接

你会发现，所有认知的基础都是记忆，如果没有记忆的能力，观察、理解、推理、想象等其他所有认知行为都不会存在，甚至不会有情绪。一个患阿尔茨海默病的人，面部甚至逐渐失去表情。人类胎儿在30周后就开始了最初的记忆，婴儿从刚生下就能分辨出母亲的声音了。

如果认知的基础是记忆，那么记忆的基础又是什么呢？你仔细想想，记忆其实就是一种关联。你在学“o”这个字母时，是把一个圆圈的图像和一个“欧”的发音关联起来。那这种关联在大脑中是如何形成的呢？

这种关联是通过我们大脑中神经元之间的连接形成的。大脑有大约1000亿个神经元，一个神经元可以从许多其他神经元接收电脉冲信号，同时也向其他神经元输出电信号。

如图1.2所示，每个神经元都能输出和接收信号。负责输出的一端叫“轴突”，负责接收的一端叫“树突”。每个神经元都有几千个树突，负责从不同的神经元接收信号。同样，每个神经元的输出信号可以传给和它相连的几千个神经元。那么这个最初的信号是从哪里来的呢？通常都来自感觉细胞，如视觉细胞、听觉细胞等。

图1.2大脑神经元和突触的结构

图片来源：[https://www.researchgate.net/figure/Generic-neurotransmitter-system\\_fig1\\_318305870](https://www.researchgate.net/figure/Generic-neurotransmitter-system_fig1_318305870)。

那神经元之间是怎么连接的呢？一个神经元的轴突和另外一个神经元的树突之间有20纳米（一根头发丝的1/2000）的小缝隙，这个缝隙叫“突触”。图1.2的右半部分就是放大的突触。它保证了两个神经元各自独立，不会粘在一起。记忆的主要奥秘就藏在这里。在这个连接的地方前一个神经元的电信号会转化成化学物质传递到下一个神经元，下一个神经元接收到化学物质后又会再转成电信号。不同的突触面积大小不同，化学物质的传递速度和量不同，因而造成有些突触是“貌合神离”，相互之间并没有电信号通过；有些则是“常来常往”，经常有信号通过。

你一定听说过俄国生理学家巴甫洛夫（Ivan Pavlov, 1849—1936）的条件反射实验。受到条件反射的启发，加拿大心理学家赫布（Donald Hebb, 1904—1985）在1949年提出了一个大胆的猜想。他认为当大脑中两个神经元同时受到刺激时，它们之间就会建立起连接，以后其中一个神经元被激发时会通过连接让另一个神经元也被激发。譬如在巴甫洛夫对狗的实验中，送食物时同时摇铃，摇铃刺激了听觉神经元，食物味道刺激了嗅觉神经元并且导致分泌唾液，听觉和视觉神经元同时受到刺激，它们之间就建立了连接，一个神经元的激发会导致另一个神经元的激发。经过多次反复，它们的连接会越来越稳定。以后即使没有送食物，狗只要听到摇铃就像闻到食物一样会分泌唾液。人也是一样，比如说一个小孩子被火烫过一次就能把“火”和“疼”联系起来。当小孩看见火时，他大脑中负责接收视觉信号的神经元被激发了，与此同时他的手感觉到烫，于是他大脑中负责接收皮肤感觉细胞的神经元也被激发了。如果看到火和感觉到疼这两件事同时发生，那么这两个神经元细胞就连通了，也就是有信号通过了。下次这个孩子见到火，马上会想到疼，因为当负责看到火的神经元被激发后，马上会把信号传给负责“疼”这种感觉的神经元，就能让小孩想到疼。刺激越强，神经元的连接就越稳固。孩子被火烫过一次手就永远记住了，再也不会去摸火；有些刺激很弱，连接就不稳固，长时间不重复就会断开。例如背英文单词，重复的刺激越多，信号的传递速度就越快。比如一个篮球运动员对飞过来的篮球的反应比普通人快很多，一个空军飞行员对飞机姿势和敌人导弹的反应都比普通人快，这些都是反复训练出来的。所谓赫布猜想，本质上是通过建立神经元之间的连接从而建立起不同事物之间的联系。后来这个猜想被科学家反复证实，就成了现在我们常说的赫布学习定律。

赫布定律揭示了记忆或者说关联的微观机制，启发了好几代计算机科学家，他们开始用电子线路模仿神经元，然后用许多电子神经元搭建越来越大的神经网络，今天这些神经网络的记忆和关联能力已经远远超过了人类，许多机器的“神迹”大都源于这种超强的记忆和关联能力。在第三章，我们会介绍为什么神经网络的超强记忆和关联能力会转化为不可思议的“超人”能力。

这些在大脑中由神经元的连接形成的关联记忆又可以分为两类：可表达的和不可表达的。

## 可表达的“明知识”

目前，脑神经科学的最新研究发现，可表达的记忆并不是对应着一组固定神经元的连接，而是大致地对应于散布在大脑皮层各处的一些连接。原因是用来表达的语言和文字只能是体验的概括和近似。这类可以用语言表达或数学公式描述的知识就是人类积累的大量“正式知识”，也可以称为“明知识”。它们记载在书籍、杂志、文章、音频等各种媒体上。

要想把某种关联表达出来，人类唯一的方法是通过语言和符号。语言和符号表达的第一个前提是要有概念。所谓概念就是某个特定的发音或符号稳定地对应于一个事物或行为。大部分的名词和动词都是这样的概念。第二个前提是每个概念都不同于其他概念，猫就是猫，狗就是狗，不能把猫叫成狗，或者把狗叫成猫，两者要能区分开。这叫“同一律”。第三个前提是猫不能同时也不是猫，黑不能同时也是白。这叫“不矛盾律”。有了这些基本前提，根据已知的事物间的关系我们就可以推导出新的知识或者论证一个决定的合理性。推理、假设、联想，这些本质上都是建立在语言之上的思维活动，没有语言就完全无法思维。所有的正常思维都要借助概念，要遵循“同一律”和“不矛盾律”。语言是人类和所有动物的最大区别。黑猩猩可以学会很多概念，譬如“我”“吃”和“香蕉”等，但无论实验人员如何训练黑猩猩，它们都无法组合出“我要吃香蕉”这样的句子。人的语言能力的本质是什么？它的生物学基础是什么？语言和自我意识是什么关系？目前这些都还不清楚。但我们知道，人类语言是不精确的，越基本的概念越不容易定义清楚，像“公平”“理性”等。人类语言中有大量含混和歧义的表达，像“今天骑车子差点滑倒，幸亏我一把把把把住了”。

英国哲学家罗素企图把语言建立在精确的逻辑基础之上，他用了几百页纸的篇幅来证明 $1+1=2$ 。德国哲学家维特根斯坦（Ludwig Wittgenstein, 1889—1951）认为人类有史以来几乎所有的哲学辩论都源于语言的模糊不清，因而没有任何意义。他认为在世界中只有事实有意义，在语言中只有那些能够判断真假的论断才能反映事实。他的结论是：我们的语言受限，因而我们的世界受限。

为什么语言的表达能力受限？用信息论的方法可以看得很清楚。我们大脑接收的环境信息量有多大？一棵树、一块石头、一条狗都包含几十MB（兆字节）甚至几十GB（千兆字节）的数据，我们的感觉接收神经元虽然大大简化了这些信息，但它们向大脑传导的信息量仍然非常大，表1.1是各个感觉器官每秒钟能向大脑传递的信息量。

表1.1人体各个感官向大脑传送信息的速率

□

资料来源：<https://www.britannica.com/science/information-theory/Physiology>。

大脑存储这些信息的方式是神经元之间的连接，大脑在存储时可能进一步简化了这些信息，但它们的信息量仍然远远大于我们语言所能表达的信息量。人类语言的最大限制是我们的舌头每秒钟只能嘟囔那么几下，最多表达几十个比特的意思。（比如读书，我们平均每分钟能读300字，每秒读5个字=40比特。）这样大脑接收和存储的信息与能用语言表达出来的信息量就有6个数量级的差别。也就是说极为丰富的世界只能用极为贫乏的语言表达。许多复杂事物和行为只能用简化了的概念和逻辑表达。这就是人类语言的基本困境。

## 只可意会的“默知识”

由于舌头翻卷速度严重受限，以神经元连接形式存在大脑中的人类知识只有极少一部分可以被表达出来。而绝大部分知识无法用语言表达，如骑马、打铁、骑自行车、琴棋书画，察言观色、待人接物、判断机会和危险等。这些知识由于无法记录，所以无法传播和积累，更无法被集中。英籍犹太裔科学家、哲学家波兰尼（Michael Polanyi, 1891—1976）称这些知识为“默会知识”或者“默知识”。波兰尼举了骑自行车的例子。如果你问每个骑自行车的人是怎么保持不倒的，回答会是“车往哪边倒，就往哪边打把”。从物理学上可以知道，当朝一个方向打把时会产生一个相反方向的离心力让车子平衡。甚至可以精确计算出车把的转弯半径应该和速度的平方成反比。但哪个骑自行车的人能够知道骑车的速度呢？即使知道谁又能精确地把转弯半径控制在速度平方的反比呢？所有骑自行车的人都是凭身体的平衡感觉左一把右一把地曲折前进。世界上大概没有一个人学骑自行车是看手册学会的，事实上也没有这样的学习手册。大部分技能类的知识都类似。

欢迎访问：电子书学习和下载网站 (<https://www.shgis.cn>)

文档名称：《暗知识：机器认知如何颠覆商业和社会》[美]王维嘉 著.epub

请登录 <https://shgis.cn/post/742.html> 下载完整文档。

手机端请扫码查看：

