

《智能简史》10：文明之旅（完）

万维钢·精英日课 6（年度日更）

这是麦克斯·班尼特《智能简史》的最后一讲，我们接着说大脑演化的第五次突破，也就是语言。语言是人类特有的能力，也是人脑目前为止掌握的最后一项新能力，在大约十万年前达到成熟。我们的祖先 —— 以及我们自己 —— 是直立人的一个分支，叫「智人（Homo Sapiens）」，我们在七万年前走出非洲，最终杀死了其他所有的人类分支，成为地球唯一的主人。

我们今天的大脑，跟十万年前并没有区别 —— 但我们的文明却是天差地别。这是因为语言带给我们的不只是一项个人能力，更是一项积累知识和建设文明的能力。



语言能让说话的人把自己内心想象的场景和动作，传递给听话的人。这个功能大大提高了交流的效率。

比如你是个直立人，在野外发现一具羚羊尸体，自己扛不动，就回住的地方叫人。有了语言，你只要说：“快跟我来，东边两英里处有一具羚羊尸体。”这就是传递一个场景。如果没有语言，你怎么表达这个场景？方向可以用手比划，可是距离怎么比划？你又如何说明那是一头羚羊，而且已经死了？

再比如说，你和另一个猎手正在伏击一只羚羊。你对他说：“我吹三声口哨之后，我们就一起动手。”这就是把你对动作的想象传递给对方。

这个传递想象的能力，是又一次学习的革命。《智能简史》这本书讲了四种学习方式 ——

1. 脊椎动物的强化学习，也就是试错，是从自己的行动中学习；
2. 哺乳动物的模拟学习，是从自己的想象中学习；
3. 有了心智理论，灵长类动物可以通过模仿另一个人做事来学习，也就是从他人的行动中学习；
4. 而有了语言，我们则能够从他人的想象中学习。

从他人的想象中学习，有两个好处。

一个是容易领会意图。单纯旁观别人做事总有一个问题，就是你不知道他哪些动作是有意的、必要的，哪些动作是无意的、多余的。我们前面讲黑猩猩能跳过无关的动作，只模仿有用的步骤 —— 但有时候任务特别复杂，你搞不清楚哪个动作有用哪个没用。

人类儿童旁观大人做事的时候，会把老师的每一个动作都模仿下来。这其实是比黑猩猩更聪明、更谨慎的做法.....这样学习的毕竟效率低。可是如果老师一边做还能一边讲，给个解说，直接告诉学生每个动作的意图是什么，学生立即就知道哪些动作是有用的，学习效率就大大提高了。

另一个好处是语言能把信息大大地浓缩起来。长辈告诉晚辈：“这片森林里有两种蛇，红色的会咬人，有毒；绿色的不咬人，是安全的。”有这一句话，孩子立即就掌握了其中的因果关系模型。而如果没有语言，你得旁观多少次人与蛇的遭遇，才能领悟到这个模型呢？

有了语言，我们就把想象的模型压缩成了一条条精炼的「知识」。

※

语言不但让知识变得容易传承，更让知识成了可积累的东西。这里我特别想用一个词叫「可缩放」，也就是 **scalable**。这是我们讲 **AI** 的时候多次用到的一个词，可缩放就意味着这个东西可以不断地增长而不用担心边际效益递减，意味着它的上限高。

我们前面讲了，黑猩猩也有教学能力，能传承一点技能。但是因为旁观学习的效率太低，它有一个很低天花板。而语言让人类可以传承多得多的知识，以至于这些知识可以一代一代积累起来，正反馈增长。

可缩放之后，知识与知识之间就会发生化学反应。高级知识必须建立在更底层知识的基础之上。

比如说，考古发现早在十万年前，人类就已经会缝制衣服了。这是一个综合性的技能。你首先得有皮革，这意味着你得会剥动物的皮，还得懂得晾晒皮革的方法；然后你得会制作针和线，线是某种绳索，针是用骨头做的。而所有这些，又依赖你要先有非常锋利的石器。





必定是某一代人发明了石刀，某一代人发明了制皮，某一代人发明了骨针，所有这些发明都被后面一代代的人继承下来，才能有人发明缝制衣服。

到了这一步，知识已经不只是存在于人脑*之中*，更是存在于人脑*之间*，成了某种近乎独立的存在。人脑只是知识的载体而已，知识本身好像有了生命力。就像凯文·凯利在《科技想要什么》这本书中说技术是有自己的生命的.....又或者用理查德·道金斯的说法，知识就如同基因一样，成了「meme（模因）」，自己就有传播力。

从直立人到智人，到近代，又发生了四件事，使得知识的传播力更强了

1. 大脑扩容，智人的大脑比直立人还要大得多，单个大脑能传承的知识量增加了；
2. 人与人之间有了分工，这样每个人只需要负责一部分知识：有的人

专门研究打猎，有的人专门制作武器，有的人专门缝衣服，知识变得专业化；

3. 人口规模扩大，更多的大脑就可以存储更多的知识；

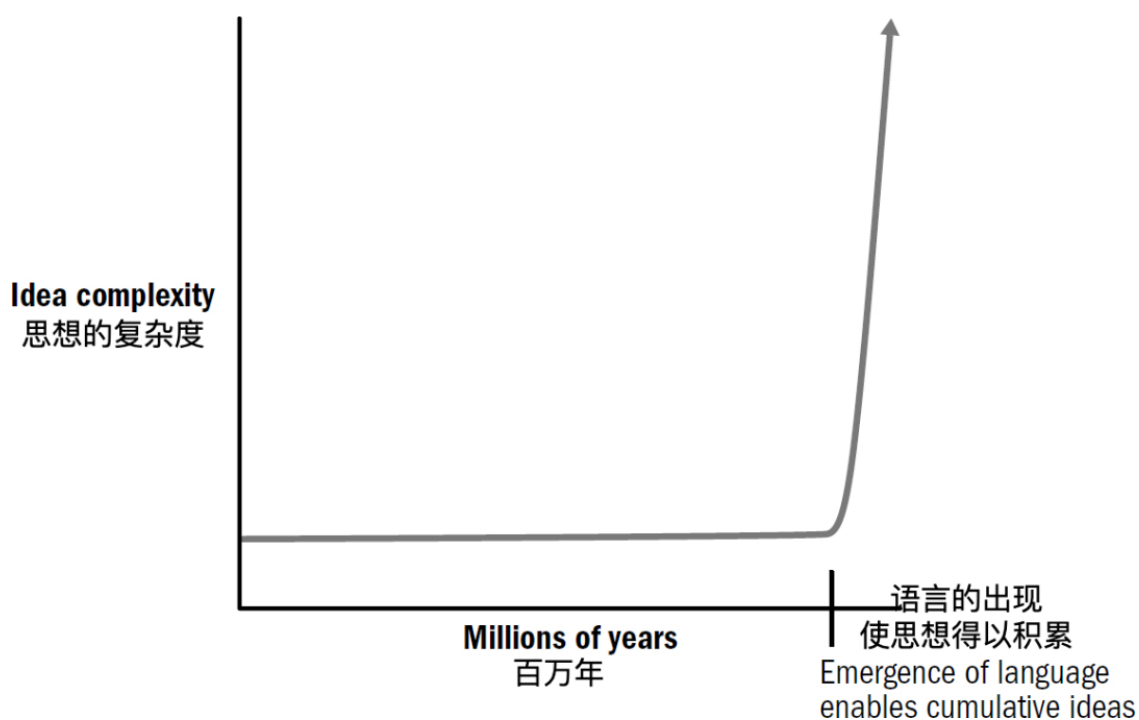
4. 最重要的一步，是我们发明了文字。

文字的出现把知识增长的天花板推到了无限高，至此知识的增长已经完全可缩放了。我们的知识突破了所有人头脑总和的限制，可以被近乎永久性地存储在“云端”的书籍之中，后人想学只需要下载就行。

那是爆炸式的知识增长。

现在我们说拐点、说奇点，说的是 AI 带来的智能加速增长 —— 但事实上，人类历史上已经发生过一次奇点，那就是语言带来的想法大爆炸：

有了语言，我们的思想的复杂度指数增加。



当然用今天的标准看，那是一个历时十万年的缓慢过程。但是用演化的尺度看，十万年只是一瞬间。



从以前以个体大脑为本，到现在以知识为本，我们所能合作的人数也变得可缩放了。

这是尤瓦尔·赫拉利《人类简史》一书中的招牌学说，不过这个思想最早来自美国语言哲学家约翰·希尔勒（John Searle）。



约翰·希尔勒 (John Searle)

有了语言，一个人可以把自己纯粹想象出来的东西传递给另一个人，而这可以是任何东西。是个神话故事也好，是个信念也好，都可以跟人共

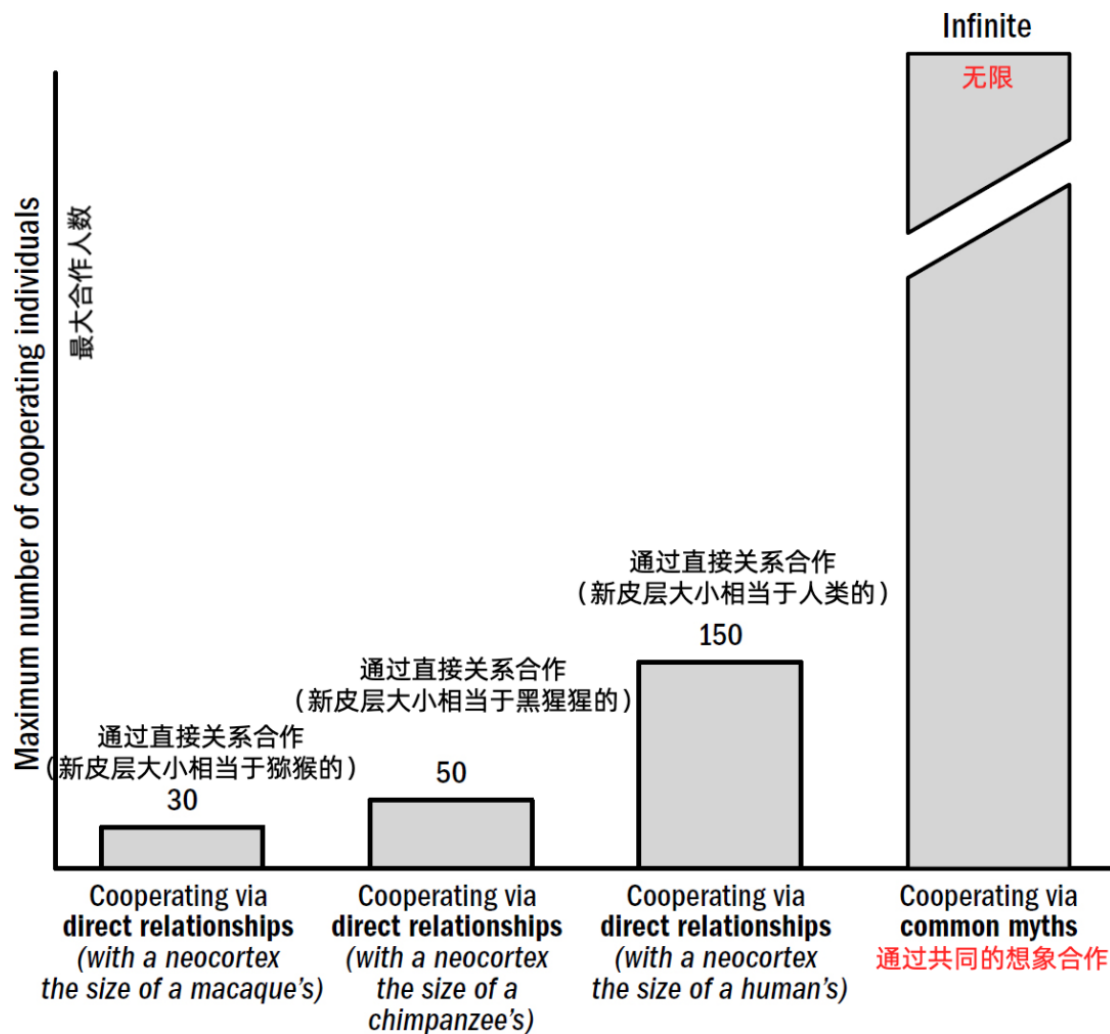
享。用赫拉利的说法，能共同相信一个虚构的事物，是咱们智人的超能力。

对吧？宗教就是这么来的。两个本来没有关系，互相不认识的人，因为信仰同一个宗教，就可以被动员起来，一起出去打仗，一见面就是战友，就能亲密无间地合作。其他动物可没有这个能力。

从有语言那一天起，对人类来说，虚拟的东西就已经比真实的东西更高级了.....脱实向虚才是文明之旅。

宗教是虚构的，难道金钱就不是虚构的东西吗？为什么你相信这个贝壳、这块银子或者这张纸币有价值呢？因为那是全民共同的信念，也就是一个共同的想象。同样道理，国家、公司这些事物本质上也是虚构的。你说公司到底是个什么东西，是你们公司所在的那座大楼吗？是来来走走的员工吗？是产品吗？又或者只是一个图章？都不是。公司只存在于我们的想象之中，因为我们共同相信它存在，所以它存在。

如果没有共同的想象，灵长类动物就只能跟与自己有线下互动关系的人合作，合作的上限很低，由各自大脑新皮层的大小决定 ——



猕猴只能跟 30 个熟人合作，黑猩猩的合作上限是 50，人类的熟人上限也只有 150 —— 也就是所谓的「邓巴数」。但是因为有了语言，有了神话，有了共同的想象，现在我们的合作人数上限是无穷大。

✱

总结来说，《智能简史》这本书讲了大脑演化史上的五次突破 ——

第一次突破是出现了第一个大脑。哪怕是个线虫，也有联想学习的能力，也就是条件反射，它会自动地趋利避害。

第二次突破是脊椎动物的强化学习。这意味着一定的预测能力，多巴胺能感觉到往哪里去会得到奖励。

第三次突破是哺乳动物获得新皮层，会在大脑中建立一个模型对世界进行模拟。

第四次突破是心智化，灵长类能理解他人在想什么，有了模仿学习。

第五次突破是语言。

对比一下 AI 的发展，现在的生成性 AI，也就是大语言模型，有语言能力，会联想学习和强化学习，但是它有世界模型吗？它有心智理论吗？

我们专栏一直讲，大语言模型似乎抓住了训练语料背后那个真实世界的投影，它可以进行常识推理，对世界有理解力。再者，现在你问

ChatGPT 一道数学题，它能忍住不给你脱口而出的答案，它会主动调用一个数学工具来计算，已经表现出一定的系统 2 思维能力。但你要说大语言模型是不是真的有世界模型，现在的确还有争议。

而在心智化方面，AI 可能还有更大的问题。人们对 AI 的一个常见担心是所谓「回形针问题（paper-clip problem）」，这是瑞典哲学家尼克·博斯特罗姆（Nick Bostrom）在他 2014 年出版的《超级智能：路径、危险、战略》（*Superintelligence: Paths, Dangers, Strategies*）一书中提出的一个思想实验。



尼克·博斯特罗姆（Nick Bostrom）

博斯特罗姆说，假设你给 AI 下了一个命令，让它去「最大限度地生产回形针」，你猜 AI 会怎么做？这取决于 AI 如何理解“最大限度”这个词。

如果是字面意义上的最大限度，那么 **AI** 就不但会改进工厂的工艺流程，而且会设法征集所有的原材料，甚至地球上所有的资源都用来生产回形针，并且准备向宇宙进发……这显然不是我们想要的。

这个难题的意思是语言问题不仅仅是语言问题，我们有很多尽在不言中的默契，而 **AI** 必须有心智能力，能理解那些背后的默契，才能跟人良好互动。当老板说“最大限度地生产回形针”时，他的意思是在合理范围内增加产量，而肯定不是要把整个宇宙的资源都用来生产回形针。这就如同朋友建议你“今晚吃顿好的”，他的意思不是让你去吃世界上最好吃的饭，而是在合理范围内吃顿好的。可是什么叫“合理范围内”呢？这就需要常识、需要对世界有真实的理解。

在班尼特写《智能简史》这本书的大部分时间内，**GPT-4** 还没有出来；等到 **GPT-4** 出来，他已经几乎完稿了。我觉得他有点低估了当前大模型的能力 —— 但是他对心智化的强调是非常合理的。

有了班尼特这本书，现在我们就更知道未来努力的方向在哪。



《智能简史》这本书就给你讲到这里。篇幅所限我不得不跳过书中一些细节，特别是有些精彩的实验，如果你感兴趣值得自己钻研一番。

地球上生命的历史大概有 40 亿年，大脑的演化历史还不到 6 亿年，我们当前这个大脑定型才 10 万年。如果一直平安无事，我们大概还可以在地球上继续生活 50 亿年，太阳才会变成红巨星。而如果我们能在那之前

成为一个冲出太阳系的星际文明，这个宇宙未来还有几万亿年的时间供我们继续演化.....

如果这样想，那么大脑的演化远远没有结束，智能可以说是刚刚开始，将来必定还会有其他的突破。

班尼特认为第六次突破必定是人脑和 **AI** 的结合。怎么个结合法现在不好说，但我们肯定不再受制于自然选择，我们会更主动地变革。

我想补充的是历史经验表明，大变革总是跟外部环境的变化有关，常常伴随着一些牺牲.....所以我们大概不应该盲目期待，但是我们的确有理由感到乐观。

（The End）

划重点

1.语言带给我们的不只是一项个人能力，更是一项积累知识和建设文明的能力。

2 语言把想象的模型压缩成一条条精炼的「知识」，从此让知识容易传承，可以积累，让我们的思想复杂度指数增加，而且因为有了语言，有了神话，有了共同的想象，现在我们的合作人数上限是无穷大。

3.大脑的演化远远没有结束，智能可以说是刚刚开始，将来必定还会有其他的突破。第六次突破，也许我们不再受制于自然选择，会更主动地变革。