

国家自然科学基金资助项目

认知科学与 广义进化论

赵南元 著



清华大学出版社

认知科学与广义进化论

赵南元 著

国家自然科学基金资助项目

清华大学出版社

(京) 新登字 158 号

图书在版编目 (CIP) 数据

认知科学与广义进化论/赵南元著. -北京: 清华大学出版社,
1994

ISBN 7-302-01446-9

I. 认… II. 赵… III. ①认识论-广义进化论②广义进化论
-认识论 IV①B017②Q111

中国版本图书馆 CIP 数据核字 (94) 第 01373 号

出版者: 清华大学出版社 (北京清华大学校内, 邮编 100084)

印刷者: 北京密云胶印厂

发行者: 新华书店总店北京科技发行所

开 本: 850×1168 1/32 印张: 11.75 字数: 303 千字

版 次: 1994 年 3 月第 1 版 1994 年 3 月第 1 次印刷

本社分类号: N·1

印 数: 0001—4000

定 价: 16.80 元

内 容 简 介

《认知科学与广义进化论》一书是著者所承担的国家自然科学基金项目的研究成果。在本书中著者创立了横跨哲学认识论、比较文化学、进化生物学、计算机科学、系统科学等多学科的高度统一的认知科学基础理论——广义进化论。并探讨了该理论在建立伦理学、美学等社会科学基础方面的作用。本书论点丰富新颖，论证严密科学，表达深入浅出，不仅适合于有关学科的教学、科研人员参考，而且适合于一般读者阅读。

序 言

本书的目标是建立认知科学的核心理论。通俗地说，就是要回答这样的问题：人的大脑（或心灵）是按照什么原理工作的？人工智能的研究曾经试图解决这个问题，但是传统人工智能在这一方面显然没有能够取得令人满意的结果。本书的第一章就是分析这个问题的困难之处，为读者提供必要的背景知识。人工智能的失误并不是技术上的失误，而是哲学层次上的失误。为了解决这个世纪性的难题，我们不得不首先从哲学层次上建立与我们的目标相适应的理论，即工程认识论。这种认识论与通常的基于本体论的认识论的区别，根植于东西方文化的深层精神结构的不同，因此我们在第二章从比较文化学的角度入手，从更深层去挖掘认识论差异的根源。认识论的差异导致了对科学本身看法的变化，因此在第二章里我们也讨论了一些与科学哲学有关的问题。

第三章是本书的核心理论——广义进化论。把广义进化论作为认知科学的基础，体现了一种被很多认知科学的研究者所忽视的事实，即人是一种动物，人是从其他动物进化而来的，人类的认知和思维能力也是在进化之中产生的。另一方面，人与机器的最大区别是人具有创造性，而为了理解创造过程，最好的方法莫过于借鉴自然界中最为壮丽的创造过程——进化。广义进化论是对达尔文进化论的发展。当我们用新的眼光观察生物界时，会产生新的观点。在第三章中我们用广义进化论的观点讨论了进化生物学的一些重要问题，展示了广义进化论的理论威力。在第四章中，我们运用广义进化论的观点，对认知和思维中的重大问题进行了分析，建立了感情和意识的具有工程可实现性的理论模型，阐

述了感情和意识在认知中的作用和机制。

创造源于进化，进化的基本机制是变异与选择，选择的根据是价值评价。因此本书在哲学层次上建立了关于价值的基础理论，并将这一精神贯彻于全书内容之中，成为全书的一条主线。为自然科学与社会科学寻找共同的理论基础，是科学对于深刻性的必然追求。由于本书建立了关于价值的深层理论、关于进化的广义理论、以及关于人类决策的认知理论，也就提供了自然科学与社会科学的共同基础。以往社会科学建立基础理论的困难主要是善恶评价的主观性、社会不断发展变化的自我表述性、以及作为社会的基本单元的人的复杂性。本书的理论对于克服这些困难是有帮助的。在第五章里我们用伦理学、美学等实例，探讨了本书的理论在社会科学和复杂巨系统理论等领域中的作用。

由于本书所涉及的学科范围很广，不可能要求读者（包括著者）在所有这些领域都是专家，而且也考虑到非专业人员阅读的可能性，所以本书的写作方针是“深入浅出、点到为止”。在文字表达上力求通俗，尽量避免数学形式的表达。对于“好读书，不求甚解”，仅以扩大知识面为目的的读者来说，本书的可读性也是很强的，足以提供茶余饭后之谈资。对于想要深入理解本书内容的人来说，需要多下些工夫才行。本书所研究的问题本身是很困难的，新概念比较多，大多数概念要通过与其他概念的相互关系才能深入理解，因此建议读者在阅读本书时可以采用多次反复阅读的方法，不清楚之处先放过去，等下次再读时，可能就容易理解了。可以说，只阅读本书的个别章节是很难理解的，任何一个章节的深入理解都需要整体的支持，这就是反复阅读的理由。此外，由于本书在有限的篇幅内容纳了庞大的理论体系，使得叙述趋于精炼，有些地方会有说明不充分之感，甚至有些跳跃。其中有些是在其他有关学科中已经研究过或正在研究的内容，读者可以参考有关文献以加深认识，更多的则是尚未深入研究的部分，可

以作为未来研究的课题。本书的贡献不仅是建立了一个理论框架，更重要的是开辟了广阔的研究领域。本书的一节，乃至一个自然段所讨论的问题，有不少可足以构成一个博士论文的题目，有待读者进行发掘。著者在清华大学开设“认知科学导论”的研究生课程，主要讲授本书的内容。本书可以作为认知科学的教科书使用。

感谢国家自然科学基金委员会对本研究项目提供的资助，为本项目创造了宽松的研究环境。

感谢日本的株式会社 *Ampere* 及社长草椰高志先生为著者在日本为期两年的研究活动提供了充足的费用和良好的工作及生活条件，在这段时间的研究中本项目取得了很大的进展。

感谢已故的常迥教授。如果没有常迥教授对本研究项目的极力支持和鼓励，著者不可能有勇气进行如此艰巨的挑战。常迥教授作为项目总负责人未能在生前见到本书的完成，是著者最大的遗憾。

感谢著者在攻读博士时的导师饭岛泰藏教授，正是他的影响使著者从一个工程技术人员转变为对基础理论抱有强烈兴趣的科学工作者。

感谢王文渊教授在本项目研究过程中与著者的充分合作。为使著者可以集中精力进行理论研究和撰写本书，他领导了本项目中所有实验研究，承担了项目组的所有日常具体工作，还接替了著者担负的行政工作，并对著者的工作不断提出积极的建议。本书能够按期完成，与王文渊教授的贡献是分不开的。

感谢李衍达教授对本书进行了审阅和出版推荐，并对本书内容提出深刻的意见，使本书进一步完善。同时感谢本书的责任编辑在本书的编辑出版工作中所付出的辛勤劳动。

感谢边肇祺教授、阎平凡教授长期以来对本项研究的全力支持和热情关怀，以及在学术方面起到的推动作用。

参与本项目研究的李小平博士、刘力夫博士、胡胜发博士、以及博士生范列湘、硕士及硕士生丁志敏、蔡奕、于婷、李庆华、黄兵、王嘉欣等，作了大量的研究工作或提出了自己的看法，为本项目作出了各自的贡献，特此致谢。

在研究过程中，林行刚教授、杨福生教授、陈霖教授、马颂德教授、迟惠生教授、马希文教授、小川英光教授、佐藤诚教授、中野馨教授、黄煦先生曾与著者就本项目有关的问题进行过有益的讨论，特此致谢。

著者所在单位的各级领导、同事及著者的家庭成员为此项研究创造了良好的环境，特此致谢。

在本书撰写过程中，钱学森教授来信对本书表示关心、肯定和支持，对本书的完成起到了巨大的促进作用，在此深表感谢。

北京清华大学自动化系 赵南元

1994年1月

目 录

序 言	(Ⅱ)
第一章 绪论	(1)
1.1 三个许诺和一则寓言	(1)
1.2 人工智能中的问题及其困难之处	(2)
1.3 认知科学的方向	(17)
第二章 工程认识论	(41)
2.1 东西方文化比较和深层精神结构	(46)
2.2 科学与形而上学	(61)
2.3 从真理到善理	(69)
2.4 认识论的逆镜模型	(96)
第三章 广义进化论	(113)
3.1 自我表述系统	(113)
3.2 进化的地形图模型	(117)
3.3 进化论的博弈论方法	(122)
3.4 广义进化论的软硬结构模型	(135)
3.5 生物的进化与进步	(151)
3.6 有性生殖的意义	(165)
3.7 广义进化范式	(178)
第四章 认知与思维	(183)

4.1	认知的阶层性	(183)
4.2	神经网络	(190)
4.3	神经网络群体	(203)
4.4	从动物到人的遗传与学习	(209)
4.5	智能模型与形象思维	(231)
4.6	感情与评价	(250)
4.7	意识与自我	(270)
4.8	创造与灵感	(291)
4.9	模式识别	(301)
4.10	人工认知主体	(304)
第五章	广义进化系统	(309)
5.1	复杂巨系统	(309)
5.2	全局观点与分离主义	(313)
5.3	伦理学的非形而上学基础	(326)
5.4	美学的情绪基础	(352)
5.5	社会的通用评价尺度——论金钱	(356)
	参考文献	(364)

第一章 绪 论

1.1 三个许诺和一则寓言

大约 40 年前，正值计算机科学的黎明时期，计算机专家许下了三个宏伟的诺言：

10 年之后

- ① 计算机可以发现有价值的新定理。
- ② 计算机可以夺得国际象棋的世界冠军。
- ③ 大部分心理学理论将采用计算机程序的形式。

40 年后的今天，这些诺言还没有一个是完全实现了的。这 40 年间，计算机硬件得到了飞速的发展，计算速度、存储容量、性能价格比、性能重量（体积）比等各项参数都按指数规律增长。如果这个增长势头能保持下去，50 年后，计算机的单位重量的计算能力和存储容量都可以超过人的大脑。但是 50 年后我们能不能用这种高能力计算机作为大脑制造出象科幻影片中出现的那种具有人的能力的机器人呢？对于这个问题，恐怕现在的人工智能研究还难以给出肯定的回答，我们所掌握的关于大脑认知过程的知识还差得很远，即使硬件条件具备，软件和制作软件的理论可能还达不到实现这一目标的要求。

对于这种现状，一位学者作了极其生动的比喻，讲了这样一个寓言：“警察问一个在路灯下找东西的人：‘你在找什么呢？’‘我在找钥匙’‘你的钥匙是丢在这儿了吗？’‘不是’‘那你为什么要在这儿找呢？’‘因为这儿比别处亮’”这个寓言本来是针对计算机视觉的研究现状而言的，但是看起来也符合整个人工智能领域

的现状。在这个领域里，研究者往往忽视计算机与人脑的不同之处，热衷于搞一些利用现成的数学工具及其他理论工具所容易解决的问题，结果是研究出的方法与实际的视觉及智能的机制相去甚远。

为了改变这种现状，需要解决的问题是：

① 钥匙到底丢在哪儿了？

② 那儿为什么暗？

③ 能不能让暗处变亮？

这三个问题换一种说法就是：

① 人工智能的根本性问题在何处？

② 解决这些问题存在什么困难？

③ 这些困难能不能克服？

以下我们先就人工智能的一些领域来分析一下上述①②两个问题，而第③个问题则是本书整个篇幅的中心内容。

1.2 人工智能中的问题及其困难之处

1. 指数爆炸

中国古代关于指数爆炸的思考见于“歧路亡羊”的故事：某人的羊丢了，他请村里的人一起去找，不久他愁眉苦脸的回来了，别人问他：“不就是一只羊么，何必如此沮丧。”他说：“不是羊的问题，我想的是，路上有岔路，岔路前头还有岔路，这样算起来村里有多少人也是不会够的”。当然在日常生活中羊是可能找到的，这是一个思考实验，显示了指数爆炸的困难。

外国也有这样的故事。一个国王非常喜欢国际象棋，就招见发明国际象棋的人问他想要什么奖赏，他在棋盘的第一个格里放了1颗麦粒，在第二个格里放2颗，第三个格放4颗，说道：按

这样放下去，直到放满棋盘的 64 个格子。大臣们计算之后发现，要拿出这么多麦子，这个小小的王国无论如何是办不到的。 2^{64} 颗麦子够现在全世界的人吃几十年！

早期人工智能研究的一个中心课题是象下棋这样的游戏。这是因为在当时计算机的发明者著名数学家冯·诺伊曼 (Von Neumann) 已经建立了关于游戏的理论基础“博弈论”，而下棋又属于博弈论中最简单的一种情况“二人零和博弈”。二人是指只有两个人参加游戏，零和是说一方胜了则另一方为负，没有两人都胜的情况，一方的所得就是对方的所失，两人加起来得失之和为零。玩零和博弈的基本原则是“极大极小准则”，即要选择一种策略力使自己的最小收益最大化。用一个简单的例子来说，假定我有 AB 两种策略，用 A 策略对方如果不注意会被我将死，但如果注意到会吃我一个车，而如果用 B 策略，则无论对方怎么下我都能吃他一个卒。按极大极小准则我应该选择 B 策略，而不应用 A 策略去下险棋，不能假定对方是个笨蛋。按照这个准则，再规定一组评价函数，例如吃一个车得 10 分，丢一个马得 -6 分等等，再把这些原则放到计算机里去，计算机就会对所有可能的走法算出分数，按极大极小原则选出最稳妥的一步棋。到此为止没什么大问题，问题出在后头。显然，只看一步棋是不行的。对手可以先给你点甜头让你上当，然后把你杀个落花流水。为了取胜必须多看几步。假如对于每个棋局可走的方法有 100 种，对于每一种走法对方也有 100 种走法，看双方一步棋要算一万次，这对计算机来说不算多，可是要看双方两步就要算一亿次，用每秒一亿次的计算机要算一秒钟，如果看三步就得三个小时，看四步得三年，这棋就下不成了。

那么能不能用制造更快的计算机的办法来解决这个问题呢？回答是否定的。因为指数规律的增长实在太快了，而计算机受物理定律的制约，不可能作到无限大和无限快，速度受光速的限制，

元件的缩小受测不准原理的限制,元件的个数受地球资源的限制。根据勃瑞姆曼极限(陈禹:《关于系统的对话》)的计算,一个由最小的元件构成的相当于地球质量的最快的计算机,从地球诞生时一直计算到现在,可以处理的信息量小于 10^{93} 位(bit,信息量的最小单位)。 10^{93} 看起来相当大,但是指数增长却可以轻而易举的突破它。例如一个围棋盘有 $19 \times 19 = 361$ 个点,每个点可以有黑子、白子、没子的三个状态,那么所有可能的棋局就是 $3^{361} > 3^{194} \approx 10^{93}$ 。看来与指数爆炸发生正面冲突就象鸡蛋碰石头。

那么能不能想出一个巧妙的算法来减少计算次数呢?这个问题涉及计算复杂性这一专门研究领域。一般认为,一类问题的计算复杂性是问题本身的性质,与算法无关。而且人的大脑也同样不能解决指数爆炸的问题。由此可以得到一个重要的结论:如果我们对大脑所作的模型遇到指数爆炸,必然说明这个模型与大脑中的实际过程相差甚远,需要重新考虑建立新的模型。换句话说,指数爆炸是提醒我们已经走错了方向的极好的警告灯。当然,指数爆炸的困难不是绝对的,如果问题的规模比较小,就能够在可以容忍的时间里得到解决。例如有一种圈叉游戏,只有 $3 \times 3 = 9$ 个格子,一方在格子里画圈,一方画叉,看谁先把自己的记号连成三个一排,这种游戏只有 $3^9 = 19683$ 种棋局,用计算机下这种棋是很容易的。只是这种游戏没什么意思。

游戏的复杂程度差别很大,最简单的游戏是有必胜法的游戏,只要游戏双方都掌握了这种法则,并完全照办,那么从游戏一开始就注定了是先走的胜还是后走的胜。如果这种必胜法不太简明扼要,人记起来比较困难,那么机器就容易胜过人了。例如三堆火柴游戏就属于有必胜法的游戏。游戏的规则是:放三堆火柴,双方轮流从任意一堆中拿走任意根火柴(至少拿一根),拿最后一根的人算输。这种游戏的解空间不大,只有三维。如果每堆火柴不超过一百根,全部解只对应于三维空间的一百万个点,可以事先

从小到大算出这些点哪些必胜，哪些必败。如果一开始的火柴数在败点上，先拿的人必输，只能等待对手出差错。如果一开始是胜点，则只要把火柴数拿到败点上交给对方，先拿的一方稳操胜券。一开局就已知胜负的游戏，当然没多大意思。发现了必胜法，等于对那种游戏判了死刑。

象棋和跳棋之类则找不到必胜法。好在其近期评价（吃子，前跳）与最终评价（将军，跳完）之间比较一致，一般来说吃子可以增加将军的机会，因此用能力较大的计算机还可以得到不错的结果，甚至能打败世界冠军，但是还显得有些死板，世界冠军多下几次就会摸出计算机的脾气，把计算机打败。而围棋则要复杂得多，落子的近期效果与远期效果未必一致，评价准则也很不明确，所以计算机下围棋的程序至今也还没有达到一段的水平。

早期的人工智能从玩游戏和证明定理出发，希望找出能解决所有问题的通用方法来，即开发“通用解题者”的程序。这一方面遇到了指数爆炸的困难，另一方面又让人感到不符合人思考的实际情况。人在下棋时显然不会考虑几亿个可能的棋局，只能借助知识在不太多的有希望的走法中选择。出于这样的认识，60年代后期人工智能开始转向对“知识”的研究，希望找到一种通用的在计算机中表现知识和利用知识去解决问题的方法。在这个时期出现了很多知识表现手法，例如声明型知识（什么是什么）、过程型知识（怎样作什么）、知识框架、剧本型知识等等，并制成了专家系统，把专家的知识放在计算机的知识库里，用这些知识解决一些专门领域的问题。这个研究方向也遇到一些共同的困难，第一是专家的丰富知识并不全是说得清楚的，有一些知识近乎直觉，说不清的知识不能形式化，也就无法放到计算机里去。与此有关的另一个困难是，这些知识表现方法都不是面向学习的，即机器不能自己学习新的知识，而真正要作出能应付实用的专家系统，需要的知识是非常大量的。这些知识由人一条条写好送到机器里去，

其工作量有时会大得惊人，特别是考虑到人的日常生活中的知识（常识），其规模之大远远超乎一般人的想象。特别是除了严密的科学领域之外，一般的知识往往包含例外情况，而不考虑例外情况的专家系统一遇到例外就会出错。但考虑了例外则又会使知识极度膨胀，明斯基（M. Minsky）曾举了一个生动的例子来描述这种情况：“鸟能飞——如果不是企鹅或鸵鸟，如果还没死，如果翅膀没断，如果没有关在鸟笼里，如果脚没铸在水泥里，如果没有遇到过足以使其不能飞的心理上的恐惧经验。”^[1]显然这些还远没包括所有的例外因素，何况例外之中还有例外，岔路上还有岔路。

上述的研究方向统称符号主义的方向。符号主义方向在学习方面存在着根本性的困难，这是因为人或机器能够直接从客观世界取得的信息都是模式信息，例如文字、图象、语音等等，而模式信息的处理又在符号主义的视野之外，要想直接从客观世界中获得知识，只处理符号是不够的。当符号主义方向的局限性日趋明显之后，和符号主义的历史一样久远而一度受到冷落的连接主义方向重新获得了重视。

早在40年代冯·诺依曼提出了计算机的基本结构之后不久，就研究过逐次处理的计算机与人脑的区别，提出了神经网络型计算机的想法。50年代末罗森布拉特（F. Rosenblatt）试制了感知机（*Perceptron*）相当于双层的神经网络，具备学习和处理模式的能力。60年代末，明斯基等人（Minsky, Papert）出版了一部名为感知机的专著，指出二层神经网络只能解决线性可分离的问题，解决不了异或（*XOR*）问题，而多层神经网络虽然能解决任意复杂的问题，却缺乏有效的学习算法。这个结论使得神经网络的研究一度冷却下来。到80年代，出现了一些关于神经网络的新算法，特别是Rumelhart和McClelland提出的误差反向传播算法（*Back Propagation*，简称BP算法），证明了多层网存在有效的学习算法，并

实际解决了包括异或问题在内的一批问题，使得关于神经网络的研究掀起了一个高潮。通过近年来的深入研究，神经网的局限性也越来越明朗化了，神经网络的学习时间随问题的复杂程度呈指数增长。这里同样存在着指数爆炸。例如如果让神经网络去通过示例学习区分质数与非质数的一般性方法是做不到的。值得注意的是，这种难度过高的问题人也解决不了，如果对一个会数数但还不懂乘除法的儿童提示大量的质数与非质数的实例，让儿童学到能判断一个未见过的数是不是质数，同样是作不到的。这使我们想起冯·诺伊曼在研究计算机与人脑的不同点时得到的重要结论：计算机可以处理逻辑深度很大的问题，而人则只能处理逻辑深度很小的问题。从这一角度上看，神经网的特性与人脑是相近的。但是如果把人脑看成是一个均匀的大神经网络，不仅不符合脑神经科学的现实，而且也会遇到数学上指数爆炸的障碍。大脑是由大量的小神经网络模块整然有序组织起来的大系统。要想知道低能力的神经网络怎样组成高能力的大脑，还必须知道更高层次的组织原则。这已经在神经网络的研究领域之外了。

符号主义与连接主义扎根于不同的哲学土壤。从认识论角度来看，符号主义属于先验论的哲学体系，连接主义则属于经验论。现在看来两者各有其片面性，但是两者简单的合取或折衷未必能产生更好的认识论，这里需要新的哲学体系。

2. 模式识别与丑小鸭定理

模式识别是人工智能的一个重要领域，其目标是把模式信息变换成便于计算机处理的符号信息。模式识别的处理对象范围很广，主要可以分为二维和一维模式两大类（三维以上的模式信息量太大，难以直接处理）。二维模式识别主要包括图象识别，文字识别，物体识别，纹理识别，景物识别，三维运动目标识别等等；一维则包括语音识别及各种波形识别。模式识别的过程大致可以

分为两步，第一步是抽取特征，第二步是根据特征来判断模式应属于哪一类。第二步的判别在理论上比较成熟，而第一步的特征抽取则全靠系统设计者的技巧。特征抽取得不好，则整个系统识别的正确率大大下降。特征抽取的困难在于缺乏统一的方法，对于不同的目的，需要抽取的特征完全不同。例如在识别手写文字时，应尽量去掉由于不同的人写字的习惯不同而造成的变形，抽取区别不同文字的特征；而如果目的在于识别这字是谁写的，那么就要求抓住完全相反的特征。

模式识别的过程就是判别一个模式属于哪一类的过程，识别的前题是分类。在识别之前人或机器内部必然已经存在一个类别系统，否则不可能判断模式属于哪一类。如果我们能够知道人脑中的类别系统是怎样生成的，或者说知道类别系统本身是按什么特征组织起来的，那么按照相同的特征去进行识别，应该是万无一失的。如果我们再进一步能知道人脑按什么原则去选择特征来建立类别系统，那么我们就找到自动决定如何抽取特征的方法，不必再依靠就事论事的技巧来设计模式识别系统。这里得到的启发是：识别 (*Recognition*) 就是再认知 (*Re-Cognition*)，要想弄清楚识别就必需先弄清认知的过程，其中包括人是在大脑中建立类别系统的。这种看法好象是把一个简单问题变成了一个更困难的问题，但是如果非此不能解决“简单”问题的话，那么这条“弯路”还是非走不可，就象三等分角和五次方程求根这种貌似简单的问题却需要高深的数学工具才能解决一样。

早在 60 年代，美籍日本学者渡边慧证明了一条著名的定理——丑小鸭定理。这条定理的主要内容是，世界上所有事物之间的相似程度都是一样的，丑小鸭与天鹅之间的差别与两只天鹅之间的差别一样大。从这条定理可以得到的一个推论是，不存在“纯客观”的分类准则，人进行分类所依据的一切准则都是主观的，而选择什么准则进行分类则纯属主观评价问题，是一个涉及到价

价值观的问题。为了解释这个看起来违反常识的定理，渡边慧举了一个例子：按生物学的分类准则，鲸是哺乳类与牛属于一类，而与鱼不同类；按产业界的分类，捕鲸业属于水产业而不属于畜牧业，鲸与鱼在一类而与牛不同类。这种分类准则的不同源于分类的目的不同。因而选择特征时所持的价值观也不同，这些都是主观因素而非客观因素。

科学是要求客观性的。从中学作理化实验开始，教师就教我们读表针和看量筒如何避免主观误差。价值观因人而异，在作科学论证时，“我就喜欢这样”是不能作为论据的。因而考虑主观因素的议论，似乎已经超出了科学研究的范围，这就是研究认知所遇到的一个重大的困难。要克服这个困难，须要重新讨论一些问题，例如客观性意味着什么？评价有没有客观标准？等等。这些问题都涉及哲学，解决这些问题需要哲学层次的分析和研究。

3. 机器翻译与解释学循环

实现机器翻译，是早期人工智能研究的一个中心课题。为此很多国家设立了庞大的研究计划，投入大量资金研究开发机器翻译系统。60年代后期，美国的一个研究机关提出了一份报告书，认为近期内开发机器翻译系统是不现实的，机器翻译一度冷却。80年代又有一部分人认为当时机器翻译的困难是由于硬件能力不足，现在有了能力很强的计算机，已有条件研究机器翻译，又形成一个机器翻译热，但时至今日，也还没有一个令人满意的机器翻译系统问世。

几十年的努力决非徒劳无功，除了可以开发出提高人的翻译效率的辅助系统之外，至少我们对于机器翻译所遇到的困难比较清楚了。机器翻译的困难包括表层的困难和深层的困难。表层困难之一是语法的例外，人所实际使用的语言（自然语言）与计算机语言不同，语法现象经常出现例外，这使得计算机中需要存储

的语法庞大得难以实现,有时甚至关于一个词也要有很多条语法。另一个表层困难是词的多义性,只要查一下词典就可以发现,自然语言中的一个词可以有多种含义,而一种意思也可以用很多词表达。在众多的语义和表达之中选择最佳者对于计算机来说相当困难。然而,人在用自己掌握的自然语言进行会话时对于这些困难甚至感觉不到。这是由于人可以理解语言的语义,而机器则很难作到这一点,这是一个深层的困难。实际上,由人来作翻译时对于不理解的东西也是无能为力的,这也就是为什么专业著作的翻译必须由懂得该专业的人来完成,而一般非专业性的翻译也要求对于两种语言各自的文化背景有足够的了解。

虽然人们很早就知道翻译只靠语法和单词对应表是不够的,必须依靠语义信息,但是关于语义的研究所获不丰。这里的困难在于,理解总是需要一定的背景知识,而如前文所述,建立一个相当于一般人所拥有的常识的知识库是非常困难的。如果一个机器系统能象人一样,从少数的知识出发,通过对语言的理解,逐步增加其知识,就有希望通过学习得到一个知识丰富的系统。为了达到这个目的,我们需要对理解的过程有更深入的理解。理解作为动词使用时,指从不理解的状态进入理解状态的过程。这个过程我们称之为“解释学的循环”。即通过理解或解释,我们可以把少数的知识扩展出一点,而扩展后的知识又使我们能理解的范围扩大,使原来不能理解的东西有一部分变为可理解的,只要实现了这个循环,就可以象数学归纳法一样无限地扩展知识。怎样在机器上实现解释学循环是一个很重要的问题,也是个困难的问题。

4. 学习与创造

通常的计算机系统不具备学习能力,我们称之为 L^0 型系统,在 L^0 型系统中追加一个学习程序,使得 L^0 型系统能在学习程序的控制之下改变其参数或结构,这个系统就获得了学习能力成为

L^1 型系统。现在常见的学习系统几乎都属于 L^1 型系统。如果希望学习系统能在学习过程中提高学习能力,应使 L^1 型系统中的学习程序本身也可以改变,这就需要追加一个控制一阶学习程序进行学习的二阶学习程序,构成 L^2 型系统。按这条路走下去,为了得到高度的学习能力,需要建立一个 L^∞ 型系统。这就会引起无限退行的困难,一个有限的物理系统装不下无穷多个学习程序。为了克服这个困难,要在有限的系统中实现无穷的学习能力,需要一个环状的结构,即可以改变自身的学习程序,或者说知道自己在干什么的系统。从这里我们可以得到一个启发,一个具有高度学习能力的系统,必须是一个有意识的系统。人脑的高度的学习能力与其意识是分不开的。而怎样建立起具有意识的系统也是一个困难的问题。

人脑与计算机的一个重要的区别在于,人脑具有创造性而计算机只能按程序行事,不敢越雷池一步。创造意味着作出不可预知的好结果。“不可预知”与“好”是两个缺一不可的因素。对于可以预见的结果,我们通常不认为是个创造,同样,对于机器的偶发故障或是掷骰子得到的数字集,虽然不可预知,但由于不满足“好”的标准,通常也不算创造。

按照这两条标准,阿希贝 (W. R. Ashby) 提出的“超稳定系统”可以算是有创造性的机器的最简单的模型 (图 1.1)。这个模型由 X 和 S 两部分组成,当 X 的状态使得输出 G 不满足条件 η 时, S 不断发出随机信号通过 U 使 X 不断变化,直到 X 的状态使 G 满足条件 η ,则 S 进入静止状态,不再发出随机信号,机器进入稳定状态。在这个系统中,随机信号是不可预见的,条件 η 则是“好”的标准。

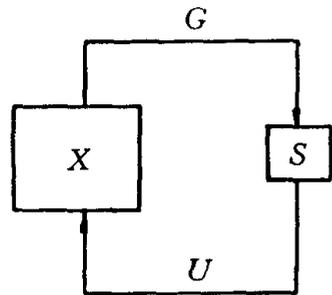


图 1.1

我们用一个具体的例子来探讨一下实现这种系统的困难在何

处。假如要用这个原理来设计一台自动作诗机（作诗当然是一种创造活动），诗中所用的汉字限于第1级和2级的6千多个字，只限于作五言绝句，每首诗20个字。作诗机中相当于S的部分每次发出1~6000的20个随机数，X按这些随机数依次选出20个汉字排成一首“诗”，X还包括一个判断是否是好诗的功能，如果不是好诗，则由G送出一个信号，让S再发下一组随机数，如果是好诗则让S暂停，把诗打印出来。

实现“作诗机”有两个显而易见的困难。第一个困难是指数爆炸， 6000^{20} 是个相当大的数，如果要作七言律诗，则已超过了勃瑞姆曼极限。第二个困难是如何让机器具有判断好诗坏诗的能力，即如何让机器富有诗意。要克服第一个困难需要一种回避指数爆炸的方法，要克服第二个困难则要有一种机器能够接受的美学理论或评价理论。在“发现有价值的新定理”这一任务中，也遇到与此相似的困难，计算机虽然可以推导出新的定理，但是，一者“岔路前头还有岔路”，二者缺乏适当的准则去自动评价一个定理是否“有价值”，是否符合作为一个定理的“美学”要求。

5. 图灵试验与感情意识之谜

图灵(A. Turing)大概可以说是最早深入地思考“机器能否思维”这一问题的人。他不仅对这一问题作出了肯定的回答，而且提出了检验机器是否能思维的具体方法——“图灵试验(Turing Test)”。图灵试验的基本想法是，两个人在相互看不见的地方通过电传打字机进行对话，如果一方的人无法区别与之对话的是人还是计算机，计算机就通过了图灵试验，可以说计算机能够思维了。图灵试验既是一种测试方法，又可以说是思维的一种定义，一种“行为主义”的或是“操作主义”的定义。关于思维的这种定义基本上是令人感到可以接受的，因为我们可以同意“思维”与外形、说话的语调无关，当一个人的外貌严重受损，发声器官破坏时思

维可以不受影响，按照一般的生活经验，我们可以通过对话对别人的思维能力作出正确的判断，而且图灵试验对于谈话的内容不作任何限制。图灵试验是相当严格的，以至于至今还没有人编出能通过图灵试验的程序。

图灵试验虽然有其令人可以接受的一面，但其行为主义基础却有很明显的缺陷。行为主义是心理学界流行过一个时期又被否定了的研究方针，其基本主张是进行心理学研究时反对一切内省和推测，只从可观测的行为出发，以保证心理学遵循科学所必须重视的客观性准则，排斥诸如意识、感情、知觉、目的、意向之类的依赖主观观察的用语。对于简单的系统，行为主义方法可能凑效，但对于复杂系统，特别是象人脑这样的复杂巨系统，行为主义显然无能为力，相同的行为背后显然可以有不同的内部状态与之对应，单靠行为来推断人脑的内部状态和过程是不现实的。举个简单的例子来说，在地上固定一根铁管子，管壁上遍布压力传感器，管子里有一个全电子化的录音设备，事先录好惨叫、喊疼、谩骂、呻吟等各种声音，只要电源一开，机器可根据所受撞击的大小、频度，次数等选择不同的声音播放出来，当我们用棒子打这个铁管时，其行为可以与感觉到疼的人维妙维肖。现在的问题是“这个机器能真的感觉到疼吗？”从行为主义的立场看，应该回答“疼”，而我们如果知道这台机器的内部构造，肯定回答“它不疼”。这里涉及一个古老的哲学问题：“子非鱼，安知鱼之乐！”如今是二十世纪了，两千多年以前的问题也该有个了结。对于机器疼不疼的问题，如果不能靠行为主义下结论，就得解决一个问题：一台机器具有什么结构或功能就可以认为它感觉到疼了呢？这个看起来极难的问题可以用进化论的方法论找到一个线索。进化论的方法论允许我们对于生物的性状提出这样的问题：这种性状为什么在生物的进化过程中不被淘汰？或者用其简化形式：这种性状对生物有什么好处？如果知道“疼”和“乐”有什么用处，就

有可能推测出“疼”和“乐”要求哪些功能，需要用什么样的结构来实现这些功能。有了这些知识，不仅可以从机器的结构上分析机器是否真的感觉到疼，而且有助于我们设计一台能觉得疼或乐的机器。这也是本书为什么把进化论作为核心理论的原因之一。

图灵曾经深入地考虑过很多关于机器思维的问题：能不能在有机的、电子的或是其他的物质基础上实现真正的智能？心灵是否不仅仅是模式？我们怎样才能区别真正的心灵和仅仅看起来象心灵的巧妙制品？自由意志是不是与唯物论者的机械论生命观不能相容？由某种规则规定的创造性的想法是否含有某种矛盾？我们的感情与知性是否属于我们自身的不同部分？机器会不会迷恋上某种新奇想法、某个人或另一台机器？机器会不会相互之间感到魅力而堕入情网？堕入情网的机器之间存在什么样的社会通用准则？机器的恋爱有无正当与不正当之分？……

霍夫施塔特(Douglas R. Hofstadter)在罗列这些问题时⁽²⁾不乏幽默感，但是这些问题无疑是非常重要的，包括图灵在内的每一个思考机器能否思维的人都无法回避这些问题。如果在1954年，年仅42岁的图灵没有由于英国的陈腐道德观念和野蛮的医学理论而身心受到极度摧残被迫害致死，而是逃到自古以来对同性恋持宽容态度的中国，这位刚过80大寿的老学者大概可以为我们作出这些问题的解答。但是图灵试验的行为主义立场对这些问题却是采取回避而不是正面解决的态度。大概经过深思熟虑之后，图灵也认为回答这些问题极为困难，因而行为主义的立场更为安全可靠。行为主义的合理性在于，我们每个人都不能进入他人的内心世界，直接观察他人是否和自己一样有意识、有感情、有知觉，只能通过他人的行为间接地判断。如果不承认这种间接判断，其结果必然导致唯我论，认为世界上只有我一个人存在，这显然是非科学的态度，因为任何科学的活动，例如发表论文，

都只有在假定他人存在时才有意义，否则写书写论文给谁看呢？

但是行为主义也有不尽合理之处。如果我偶然踩了猫脚，听到一声尖叫，我对于猫会感觉到疼这一点毫不怀疑，但是如果这是一只制作精巧的玩具猫，我会有完全相反的结论。把对待他人的态度直接推广到机器身上有其不自然的一面。在人工智能学界不乏这样的例子，某个人工智能程序在进行演示时表现非常出色，但是一旦人们了解其内部工作原理，总是异口同声地说：“这不算智能”。弄得一些人工智能专家也抱怨说：“不知道的都是智能，一旦会作了就都不是智能了。”这里有个奇妙的现象，谁也不知道真正的智能程序是个什么样子，但是谁都能判断某个程序不是真正的智能。这种神秘的直觉从何而来呢？这个貌似困难的问题其实非常简单，虽然谁也不知道真正的智能程序是个什么样子（因为世界上还没有这样的程序），但是每个人都拥有这样一个程序（自己的大脑），谁也不知道自己的大脑是怎样运转的，但是每个人都能观察自己大脑的某些运行结果。我们只要核对智能程序的中间过程与自己大脑运行的可观察到的部分是否一致，就可以判断某个程序是否真的有智能。要知道烤鸭的滋味只要吃过就行，要判断吃的是不是正宗并不需要知道烤鸭的全部工艺过程。

行为主义巧妙地回避了意识和感情这样的困难问题，但是也因此就拒绝了研究意识和感情对于研究智能的帮助。这样作得不偿失，因为机器的好处恰恰在于我们容易观察其内部，而机器又缺乏像猫一样和人类在进化上的同宗作为对类似行为作类似解释的理由。每个人都知道自己有意识、有感情、有知觉、有目的，一个自称智能的程序或是关于智能的理论不对这些现象作出说明是不能令人信服的。哪怕你说这些都是误解或幻觉，也得说明产生这些幻觉的理由和机制。意识、感情、知觉、目的等等的困难问题，如同疼不疼的问题一样，也可以依靠进化论的方法论，通过

研究这些现象在智能活动中的作用来给以解决，同时也就解决了智能的问题。

6. 进化论与麻将原理

听说美国有一股势力主张在学校讲授进化论的同时也应该讲授神创论，有一定文化水平的中国人对此大概都会感到惊奇。但是我却有幸从企图向我传教的人那里借到一本反对进化论的书，其中反对进化论的理由竟然是我们为之烦恼已久的指数爆炸，着实令人茅塞顿开。反进化论的理由是这样的，生物的基因中包含着极大的信息量，例如人类基因信息量可以抵上一座大规模的图书馆，靠随机的方法去生成如此之大的有序信息所需要的时间显然远远超过宇宙的寿命。这个计算是不错的，一首七言律诗就够吓人的了，何况一个图书馆。但是人毕竟出现了，这说明进化的机制中包含着回避指数爆炸的算法。进化是自然界中自发实现的最富成果的创造活动，对于创造的两大问题指数爆炸和评价，都提供了解决的指南，没有人参与的进化，其评价不可能是主观的，只能是客观的。这是本书把进化论作为核心理论的主要原因。

进化用以回避指数爆炸的原理我们称之为“麻将原理”。如果我们的麻将规则要求只发牌不打牌，发好牌后如果没“和”就又洗牌重发，那么我们可能打一辈子麻将也没“和”过一次，问一问常打麻将的人就可以知道，刚发到手的牌就是“和”的概率是非常小的，其原因就是指数爆炸。但是只要打牌，不管开始手中的牌有多坏，总能在可以容忍的时间内让牌“和”了，虽然我们摸牌时仍是随机的，但通过逐步的选择，淘汰不要的牌，可以逐步逼近“和”的状态。这相当于进化中的变异与选择原理。

按照进化论的麻将原理，我们可以设计第二个版本的作诗机。

这种作诗机不是一次发生 20 个字，而是每次随机发生一个字，经过选择认为这个字作为诗的第一个字很好就固定下来，不好就弃之重选，第一个字固定后按同样方法选第二个字，直到 20 个字完成，每个字选择时都要顾及已固定的字的影响。第二版作诗机仍存在评价与选择的困难，但指数爆炸已经不存在了，选择的次数充其量不过 $6000 \times 20 = 120000$ ，这是可以容忍的。字集的总字数为 M ，诗的长度为 N 时，第一版作诗机的计算时间为 M^N ，第二版为 MN ，与长度成线性关系。

当然，指数爆炸的问题并不都是可以用这种方法对付的，进化所包含的原理也并不仅仅限于一个麻将原理，否则这本书也就没有必要再写下去了。现有的所谓“遗传算法”是应用麻将原理的一个实例，但是这个算法并没有显示出明显的优越性。为了弄清进化回避指数爆炸的更丰富的内容，我们需要对进化的过程进行更深入的研究。

1.3 认知科学的方向

1. 从人工智能到认知科学

语言哲学家塞尔 (John R. Searle) 把人工智能分为“强”的和“弱”的两种^[3]。“近年来，计算机对人类的认知能力进行了模拟，这种模拟的心理学和哲学意义是什么呢？为了回答这个问题，我觉得有必要先区分几个不同的概念，它们是“强的”、“弱的”和“谨慎的”人工智能。根据弱人工智能这个概念，计算机在研究心灵方面的主要价值是提供了十分得力的工具。比方说，有了计算机，我们能够更严密、更精确地提出假设，并对之进行测试。但是，根据强人工智能这个概念，不仅可以说，计算机是研究心灵的工具，而且具备恰当程序的计算机本身就是心灵。因为一旦计

计算机有了正确的程序，我们就能说，它实际上就有了理解力以及其他的认识状态。在强人工智能里，由于输入了程序的计算机具有认识状态，因此，程序不仅仅只是帮助我们测试心理学解释的工具，程序本身就是解释。”

我当然认为现存的任何程序（SHRDLU 之类）都还不满足强人工智能的要求，不过更反对塞尔认为强人工智能不可能实现的观点。关于这一点的争论已经进行得够充分的了，本书没有必要重复。这里只是想借用强人工智能和弱人工智能的概念来说明人工智能的两个不同的发展方向。人工智能经过 40 年来的研究，其局限性和困难之处似乎已经比较明朗了。面对这些困难，一种想法是，何必费大力气去追求解开大脑认知之谜这种很可能是虚幻的目标呢？人工智能几十年的研究虽然离强人工智能的要求还差得远，但毕竟不断产生着新的成果，有些成果还颇有应用价值，即使放弃说明大脑认知过程的要求，可干的事还是很多的。发明新的智能算法，开发智能的软硬件环境，研制智能应用系统，遇到计算机犯傻的时候，由人介入一下就是了。为此需要研究高质量的人机接口，用不着让机器和人一样，该人干的事让人干，机器只干机器能干的事就行了。由此就产生了我们所说的弱人工智能的研究方向。还是在路灯底下找钥匙比较稳妥，即使找不着当初丢的那把，找到别的钥匙或许能打开别的宝藏之门。

面对困难的另一种想法是，他山之石可以攻玉。既然传统人工智能困难重重，单靠传统人工智能的方法不能对大脑认知过程作出令人满意的解释，那么我们就应该再把眼界放宽一些，从有关的其他学科汲取营养，综合各学科中的理论和方法，逐步揭开大脑认知之谜。维纳（N. Wiener）在 40 年代创立控制论（*Cybernetics*）时所用的方法是一个很好的范例，各学科的科学家人聚在一起神侃，能促使人们找出一个横贯所有学科的共同原理，在控制论中这个核心理论就是负反馈，时至今日，反馈的概念已经渗透到

社会生活的各个领域，“反馈”这个词也由一个专业词汇变成了常识范围的用语。控制论的成功显示了边缘学科或“组合学科”的力量，人工智能本来也可以看成是计算机科学与心理学的一种组合，把这种组合进一步扩大，就形成了现在的认知科学。这个方向是主张在丢钥匙的地方找钥匙，如果太黑，就造个手电筒。

认知科学是一个年青的学科，它是如此年青，以至于说它年青，不如说它是个婴儿更为恰当。认知科学这个词汇首次出现于公开发行人物，可能是在 1975 年 D. G. Bobrow 和 A. Collins 编著的《*Representation and Understanding: Studies in Cognitive Science*》Academic Press (表现和理解：认知科学的研究) 一书中。1977 年 Ablex Pub. Co. 的学术杂志《*Cognitive Science*》(认知科学) 创刊，使得认知科学这个词得到了“公民权”。1979 年夏天在加利福尼亚大学圣地亚哥分校召开了第一届认知科学会议，会议主持人诺尔曼 (D. A. Norman) 在会议上所作的报告《认知科学的 12 个主题》为认知科学的研究选择了战略目标，成为认知科学的纲领性文献。由于认知科学涉及众多学科，而这些学科之间缺乏共同语言，认知科学需要一个类似于控制论中的负反馈这样的核心理论，本书的目标就是建立这样一个核心理论。

2. 六个学科？

心理学家加德纳 (H. Gardner) 列出了与认知科学有密切关系的六个学科⁽⁴⁾，它们是：

哲学，心理学，语言学，
人类学，人工智能，神经科学

这六个学科的提法出自对认知科学作出了巨大财政支持的苏龙财团委托下于 1978 年由 12 名知名学者写成的报告书 SOAP。但由于多数学者对此报告书持否定意见，这个报告书未能公开出版。在苏龙财团的财政支持下，认知科学的名称不径而走，不止

一个学者幽默地说：“有一天早上一觉醒来，发现我从生下来就是一个认知科学家。”

我们需要分析一下，这六个学科在认知科学中起什么作用。首先是哲学。在讨论哲学与认知科学的特殊关系之际，也得弄清哲学与一般科学的关系。关于哲学与科学的关系大体有三种说法，即哲学是科学的审判官，哲学是科学的牺牲者，哲学是科学的工具。从历史上看，哲学作为科学的审判官的地位是逐步衰落的。在现代科学的结论与哲学发生冲突时，作出让步的通常是哲学而不是科学。但是任何一种科学的方法论或是对于科学结果的解释，都不免依存于某种哲学背景，比较深入的科学争论，都不免进入哲学层次，科学学派之间的分歧，往往起源于其哲学背景的分歧。某一种科学学说的成败，往往也暗示其背景哲学的成败，科学中遇到的困难，常常也是哲学的困难。从这个角度看，科学好象是哲学的实验室，与其说哲学是科学的审判官，不如说科学是哲学的验证者。这也就是为什么人工智能学者对于来自哲学家的批评或是不屑一顾，或是反唇相讥，认为不过是一些外行话，而同时又面无表情地自称是符号主义者，连接主义者，行为主义者，神经达尔文主义者，计算主义者。

古代哲学与科学并没有严格的分界，哲学家同时也是科学家，哲学是关于心灵和宇宙的全部学问，甚至宇宙与心灵也没有完全分家，处于泛灵论的混沌状态。到了现代，科学逐步从哲学中分化出来，哲学的宇宙论被天体物理学接管，哲学的原子论被原子物理学代替，哲学就象一颗洋葱，正在被一层一层剥去外皮，对于精神世界的反思可以说是哲学洋葱的最后一层皮，如果这一层皮也被认知科学剥掉了，哲学还剩下什么呢？从这个角度看，哲学似乎是科学的牺牲品。但事情也不尽如此，科学虽然占领了哲学原有的地盘，但同时又为哲学提出了新的问题，创出了新的活动领地，物理学中关于量子力学的哲学思考就是一个很好的例子。

科学是旧哲学的刽子手，同时又是新哲学的产婆。与其说哲学是科学的牺牲品，不如说科学是哲学的推进剂。即使到了可以用物理的语言来表达所有的化学现象时，化学的语言也不会消失，化学的方法依然存在。科学，特别是认知科学的发展更加需要哲学层次的讨论，哲学不会被消灭，但必然会改变自己的面貌。

哲学为科学提供原形，提供方法，也提供准则，从这一点看，哲学是科学的工具。反之科学也为哲学提供问题，提供事实，提供验证，科学也是哲学的工具。任何科学都与哲学有着千丝万缕的联系，但是把哲学公然列入自身的组成部分的学科，除了认知科学之外还没有先例。认知科学与哲学紧密相关，互为工具，互为研究对象。成熟的认知科学与成熟的认识论，可能仅仅是同一个学说在不同层次上的表现而已。

心理学、语言学和人类学是认知科学中比较靠上层的学科，如果把人脑看成一个黑盒子，这些上层学科可以给我们提供这个黑盒子的输入与输出的关系。如果这个黑盒子足够简单，从输出输入的关系很容易推测出其内部功能和结构。遗憾的是人脑这个黑盒子复杂之极，这种推测也就极为困难。另一层的困难是，每个人的大脑存在着先天和后天的差异，迫使上述学科的研究不得不使用统计学手段，而这种手段又常常会引进主观武断和片面的因素。如果我们已经掌握了大脑认知的机制，这些上层学科可以对这些机制提供很好的验证方法，反之由心理学现象去推测大脑的机制则是很难成功的。这种困难与指数爆炸有关，后文还会触及。人工智能可以看作心理学与计算机科学结合的产物，人工智能中产生的困难实际上来源于自上而下分析复杂系统的困难。在数学领域中这种困难相当于反问题的困难。

从脑神经科学的方向接近认知科学是一个自下而上的方法。这种方法的困难在于，从单个神经元或小规模的神经网络到大脑之间存在相当多的层次，每个层次都有其自身的原理，而这些原理

并不能由神经元或神经网的性质直接推测出来。由神经网络走向认知，相当于建造一个跨度太大的桥，中间不加些桥墩，这项工作是难以完成的。

对于揭开大脑认知之谜可以作一个浅显的比喻，就象一个对电视机的原理一无所知的人想要弄清电视机是怎样工作的，他可以用各种方法进行研究。自上而下的方法是，他可以每天看一个电视剧，可以操作电视机上各种按键和旋钮并观察电视机的反应，也可以从电视机的天线端子送入一些信号看电视机出现什么现象。看电视剧的方法可以使他成为一个剧作家但对理解电视机的原理毫无帮助。操作的方法可以使他学会怎样使用电视机，但并不能对理解电视机的原理有什么启发。送信号的方法取决于他对电视机的理解，如果一无所知，他不知道该对电视机送什么信号，在错误理论指导下，这种方法也可能对错误理论给以证实，但他也会发现这种理论对于说明电视机的其他现象是不充分的。自下而上的研究方法是把电视机拆开，从研究晶体管的工作原理开始，画出晶体管的伏安特性曲线，或者用万用电表测量各点电压。沿这条路去研究电视机原理的困难也是显而易见的。为了理解电视机的原理，我们必须先知道一些低层次的原理：电磁场和电磁波理论、调制解调技术、宽频带放大技术、脉冲与同步技术、电子发光技术、电子束偏转技术、扫描技术、高压发生技术等等，有了这些基础知识，不仅容易理解电视机的原理，甚至可以发明出电视机来。按照我们的观点，正如识别是一种再认知一样，理解是一种再发明。

那么我们应该到哪里去寻找理解大脑认知过程所需要的低层次原理呢？值得庆幸的是，人是由动物进化而来的，现存的各种不同水平的动物可以为我们提供丰富的低层次原理，而进化论又可以使我们有充分的理由认为人类从动物界继承了大量的精神遗产。因此，照我的看法，应该把行为生物学与进化生物学算作与

认知科学密切相关的学科。行为生物学的有效之处在于，研究动物的行为比研究人的心理要单纯得多，不容易受到文化的干扰，而且不同进化阶段的动物又能提供不同复杂程度的样本，使我们能够获得不同水平的知识。而进化生物学不仅能够提供生物智能的纵向发展图式，而且能够提供创造的良好模型。虽然用进化的观点研究认知已经有不少先驱者，但是目前这种观点并未成为主流。除了由于这个研究方向发展不成熟，存在理论上的缺陷之外，还存在着一种感情上的障碍，一种“笛卡尔情结”。笛卡尔认为人是有灵魂的，而动物只是没有灵魂的自动机器。研究认知的人的注意力往往集中于人的理性的一面，而动物恰恰欠缺这种理性，因此人们总是倾向于忽视动物学的启发，热衷于研究“人之所以异于禽兽者”。但是要想知道“人之所以异于禽兽者”，不知道“人之所以同于禽兽者”是不行的。反之亦然，要想弄清“人之所以同于机器者”先得研究“人之所以异于机器者”，如果我们真弄清楚了人和机器的不同点，就可以把这些机器欠缺的部分加到机器上，使之达到人的性能。

冯友兰在《中国哲学史新编》的全书绪论中说：“哲学是人类精神的反思。所谓反思就是人类精神反过来以自己为对象而思之。人类的精神生活的主要部分是认识，所以也可以说，哲学是对于认识的认识。对于认识的认识，就是认识反过来以自己为对象而认识之，这就是认识的反思。”这话颇有道理，与认知科学的精神一致。但是如果有一百个哲人，就会有一百种反思，我们应该接受哪一种呢？我们怎样判断这反思是成功的还是失败的呢？按照认知科学的标准，我们对于认识的认识如果是充分的，那么这种认识应该能够指导我们制作一台能进行认识的机器，(或者找到充分的理由说明制造这种机器是不可能的)，否则不能认为对认识的认识是成功的。这是对认识的认识的高标准要求，而计算机为我们提供了实现的可能性，这就是“人是机器”这句话的含义。计

计算机使认知科学成为可能的。这就是计算机科学与认知科学之间的密切关系。

3. 十二个主题

诺尔曼为认知科学提出了以下的 12 个主题：^[5]

信念系统	<i>Belief systems</i>
意识	<i>Consciousness</i>
成长	<i>Development</i>
感情	<i>Emotion</i>
相互作用	<i>Interraction</i>
语言	<i>Language</i>
学习	<i>Learning</i>
记忆	<i>Memory</i>
知觉	<i>Perception</i>
行为实行	<i>Performance</i>
技能	<i>Skill</i>
思考	<i>Thought</i>

这 12 个主题体现了认知科学的研究方向，也是本书的主要目标。为了使读者对这十二个主题初步有所了解，下面对各个问题作简单的说明，其中有引号的部分是摘自诺尔曼的说明，没有引号的部分是作者加的。

① 信念系统

“把信念系统放在第一个是由于它按字母顺序偶然排在第一，但是按别的标准来选择，信念系统也有充分的价值放在第一位。因为信念系统正好位于认知科学的传统领域——知识的研究和人类

学者、社会心理学者、社会学者这些现实世界中人的相互作用的研究者的领域相重合的部分。不用信念系统这个词而改用‘知识’或‘世界知识’的称呼可能更容易为读者所接受。但是我之所以有意避免这种称呼,是因为我想强调由几个其他范畴的知识:文化、信念、包括各种世界知识的合并融合这一点。我们一生之中获得大量的知识,获得的知识其后又对我们与他人、与环境的相互作用施加影响,甚至在某种程度上决定我们内部的信息处理。这是我对这个问题的基本想法。人类学与社会学的一个重要研究领域,就是考察这种信念的结构。”

“文化的知识作为广义的知识的特殊的子集合,世代相传,在学校或家庭内传授,或是(更经常的是)通过经验而非传授获得。服从关系,社会的相互关系,社会集团的上下级关系,会话的规则,表达社会尊敬的方式这些相互作用的模式,还有其他各种模式和形式都包括在文化之中。环境的物理形态也因文化而改变,包括建筑物,道路,运输以及所有的技术风格。”

“但是谁都能一眼看出信念系统包含比文化相互作用更多的内容,也就是说还要涉及记忆与思考方法的规则。如果读者相信人的思考和记忆是结合现实经验进行的,上述想法的可信度会提高。现在我们假定逻辑推理通常是按如下的方法进行:先对给定的问题由具体的模拟作出内心模型,然后靠经验的引导得到具体的模拟问题的解,最后把结果按给定问题的要求重新解释。如果这个假定是正确的,只能认为信念系统具有为我们的思考提供主要基础的起决定作用的重要性。”

“相同的议论不仅适用于思考过程,而且对于记忆、知觉、解决问题、理解文章以及对话、法律辩论等等等等可以无限的适用。很多读者都知道由精神异常而引起的妄想型信念系统。”

“所以,信念系统不仅自身是引人入胜的研究对象,而且作为其他多数认知活动的决定性因素是极为重要的。目前对这种结构

进行形式化分析的工具还不成熟。倾向于认知的人类学者和社会学者对于情节、图式、说话的语法等表现的问题进行着大量的讨论。我自己也一贯强调表现的研究的重要性和关于结构化记忆单元（图式）的研究所具有的效果，当然对这个研究方向抱有很大的兴趣。但是现在最重要的问题还不是表现的问题。在此之前我们有必要弄清楚信念系统这个词到底意味着什么，明确信念系统对于个人和文化集团具有的意义，确定它对于认知活动所带来的影响。我自己进入这个领域，得知自己的‘纯粹的’逻辑推理、记忆的过程、社会的相互作用在多大程度上受到自身隐蔽的信念结构的影响时，受到了强烈的心理冲击。我感到，将来我可能发现这种影响比现在我所知道的更强，而决不会是更弱。”

诺尔曼所说的信念系统的概念，类似于我们常说的“思维定势”或是“思维定势的集合与结构”。对此我有两点补充。第一，我们应该把信念系统中由基因决定的部分和由文化决定的部分加以区别而不是象诺尔曼那样混为一谈。第二是应该用动态的眼光，主要不是分析信念系统的静态结构，而是分析在学习过程中信念系统怎样被建立起来。这样才能使信念系统这个复杂的问题简单化。

② 意识

“William James 在 1980 年说过：‘谁都知道注意是怎么回事’。我在关于自己的注意的演讲和讲课中也反复说过同样的意见。但这是个谎言，是个弥天大谎。其实我们并不知道注意是什么，这首先是由于我们对于意识一无所知。以往关于注意的研究只涉及注意现象的极为有限的部分。按照 James 的说法，注意是这样一种东西：

‘注意是几个以明确而生动的形式同时被认为可能的对象或思考的流程之一占领心灵的现象。意识的聚焦、集中涉及其本质。它具有为了有效处理某些信息而放弃其他信息的含义。’

对我来说，在意识的标题之下包括有意识与无意识思考的问题、自我意识、注意、认知的控制机制、意图的形成等各个问题。还包括意识现象学的各种状态的问题。催眠是否具有很强的作用，是否可能成为探究意识的强有力的潜在工具，其本质几乎一点也没弄明白，也不能说已经进行了充分的研究。”

“关于意识用不着对已经着手于认知研究的人说三道四。但是，在我们的领域之中意识就象一个后娘养的孩子微妙地位，对于其重要性谁也不表示异议，但在实际的研究或理论之中几乎未被触及。这种相对的忽视有其充足的理由。意识是研究起来最困难的课题，对于意识来说，收集实验科学所要求的确切的、有意义的证据是很难的。我们实际上对于意识的真正性质、意识所实现的功能、无意识的性质几乎一无所知。我们直到最近才开始对于意识现象学，各种意识状态，各种现象学的经验有了一点皮毛的了解。（但是这些发现大部分不是来自于心理学或哲学这种传统的研究领域）”

“而且就是对于刚才引用 James 的话所揭示的事实，我们至今也还处于无法理解的状态。为了能够理解这个事实，我们有必要比现在更进一步理解心灵的一般功能、为什么可能同时发生几个思考的流程、有意识与无意识的信息处理的不同点，以及排斥其他而把注意力集中于某个思考流程到底是怎么回事。是什么——自己之中的什么东西——掌管着注意力的集中，而那些被除外的思考流程到底怎么样了？（我相信这些思考流程之中至少有一部分作为无意识的信息处理在沉默中不受干扰地继续潜行，其后发生中断，把意识的注意重新拉向自己。）而重要的是，有意识的注意到底是什么？无意识的注意是可能的吗？是什么——也就是自己之中的谁——经验着有意识的注意过程，产生的结果？”

“这些现象之中有一部分看来可以作为认知生物所应具备的各种性质的直接推论。复杂的环境不允许生物奢侈地把全力投入

到解决某一个具体问题。生物必须能够同时关照几件事，必须由环境事件进行数据驱动，必须尽可能利用周围的一切条件，并且必须能够时刻准备着从预想不到的危险之中立即脱身。我们为自己规定的工作时常是长期而复杂的，很多是不能一下子解决掉的那种问题。我们的脑袋里的认知资源是有限的，因此必须尽可能有效地进行分配。我们的认知不能完全成为数据驱动型的，因为如果那样的话，我们的注意会完全被感觉系统提供的滔滔不绝的信息所占领。我们为了能够在各个时刻把注意力集中于最重要（或是最有趣）的对象，必须能够在思考中排除可以排除的对象。”

“为了实现这种功能，可以认为无意识的信息处理完成了它的功能。不管意识具有哪些特别的功能，并非心灵中的一切过程都需要意识。（在意图的生成和监视有意图行为的实行中意识是很重要的。特别需要意识的是在应该实行的行为还没有作为良好的子程序得以确立的场合。这一点我在别的地方讨论过。）关于这一点还要在关于技能的项目中加以讨论。”

“关于有意识与无意识的信息处理各自的作用有一些提供管窥之见的信息源。催眠体验是其中的一种方法，可以在实验室中进行某种程度的精密研究。不时有一些关于无意识的解决问题或记忆检索的体验或报告，而且留下了进行实验探究的可能性。关于注意的研究当然还有别的强有力的研究手段，一直成为活跃的研究领域。对于说错话、做错事等等的研究，是关于无意识的信息处理与有意识的信息处理、思考、动机、意图之间关系的另一个信息源。记述各种经验的文献可能是有用的。当然在这种场合，必须注意仔细地对经验的记述和经验的解释加以区别。这是因为经验者自身同时又作为外部观察者，无法保证二者能够妥善分离按不同的方式行动。”

意识将是本书重点讨论的课题。

③ 成长

“儿童决不是小型的成年人。儿童并不是这样一种存在：仅仅缺乏经验、体魄和知识，只要把大人脑袋里的东西塞到他的脑袋里去就能成为大人。作为成年人，我们具备范围很广的技能、难以估量的详细的特殊知识，以及建立起了坚固的信念系统。我们成年人不仅仅是放大的儿童，我们是另一种存在。”

“成长的研究方法已经比较成熟，现在已不处于需要启发和建议的状态。（但是关于成长的研究集中于诞生后几年的时期，对于青年期、壮年期和老年期的研究还很不够。）但是在关于成年人认知的研究之中，似乎有一种不成文的约定，认为只要弄清楚成年人的认知，就可以把儿童仅仅看成向大人前进过程中的各种中间阶段上存在的东西。这也可能是对的。但是另一方面，成年人所具备的复杂性和经验，最终把某种特性的存在覆盖隐蔽起来是完全可能的。已经实现自动化的行动，把那种行动背后所具有的结构压进意识表层之下的无法检测的无意识过程之中隐藏起来。从而使已经建成的信念系统和知识系统不让我们看到其全部内容。”

“对于大部分认知过程来说，可能通过成长周期进行研究是最好的方法。研究成长的历史可能是到达更好地理解成年人的必经之路。某些生物的个体长到成年需要花费很长时间。人要花费毕生时间学习新的概念。学习语言要花费出生后十几年的时间，而且学习新的词汇则是没有终结的。人从根本上是不断学习持续成长的动物。忽视人类行动的这个侧面而只是静态的把握认知，就等于失去了理解人类认知的最重要的钥匙。”

④ 感情

“在认知研究之中感情到底扮演了什么角色？通常我们把感情的问题交给诗人、剧作家和小说家去处理。作为一个普通人，我们欣赏美术作品，从音乐中享受到喜悦，我们争斗、发怒、高兴、悲伤、沉浸于幸福之中。但是作为精神现象的研究者，我们对于为什么如此或怎样如此却是完全无知的。”

“感情，它到底是原始警戒系统的残余，还是在人类达到顶峰的高度发达的一群状态？对此我已经在进行着支持后一种看法的议论。所以在此读者只要再次回忆起那个问题即可。”

“关于感情又如何呢？有了认知的功能，感情之类是不是就不再需要了？我们之中多数人也包括我自己可能都愿意相信这一点。关于认知功能现今的学说——不管是哪个学科领域——看起来都是纯粹理性的理论。与感情有关的可能是别的什么东西，感情看起来很可能是关于人应该作什么的一种准则，它是远古时代在进化之中残留至今的，与现在的准则大相庭径。但是，小说家、剧作家、临床心理学家和精神科医生不这样看。如果我的想法正确，也就是说如果认为认知系统服务于中央控制系统，而纯粹认知是接枝于生物学有机体的人工的存在，那么不得不认为感情在人类行动之中扮演着决定性的角色。”

“感情的研究是重要的领域，已经有很多重要发现，对认知研究给以很大启发。不能忽视我们所接受的生物学遗产，因此不能忽视感情的各种状态。Geschwind 在他的论文中告诉了我们关于被认为对于感情很重要的神经组织和被认为对于记忆很重要的神经组织之间存在的密切关系。实际上关于状态依存型记忆的取出存在着实验的证据，因此可以说我们对于与当时心灵状态在感情内容上一致的事件最容易回想起来。也就是说悲伤的往事容易在悲伤时想起，高兴的往事容易在高兴时想起。Geschwind 暗示某重神经学的控制机构接受双重的活化，一个来自下方即感情，一个来自上方即理智。我们服从感情的信号微笑、哭泣、大笑。但另一方面，按照理智的愿望或是接受语言的指令试图作出同样的行为时，不仅不能作出真的感情，连相同的肌肉运动也作不出来。几乎所有的观察者都能辨别真的行为和假的行为。”

诺尔曼对于感情在认知中的重要性仍然估计不足，如果感情与理智发生冲突，没有热爱理性这种感情作为后盾，理智是注定

会失败的。有时我甚至觉得，理智仅仅是感情的一种委婉的表现。

⑤ 相互作用

“人是社会的生物。我们的智能不是孤立地工作，而是一般都在共存状态下工作。我们和他人具有相互作用，通过文化传递知识。我们通过社会相互作用，通过利用环境，通过人造物品来补充我们的智能。人造物品包括读写的手段（纸张、笔、印刷机）、运输系统、长距离通讯手段（邮政、电报、电话、电视）商用及其他生活用机械、计算机等等。其结果是社会的相互作用已经成为我们行动的一个本质的方面。在某种意义上我们的智能已经部分地外化，可以说在人工物品中储存着不亚于我们头脑中的知识。”

离开文化的作用则无法解释人的认知，相互作用是认知的重要组成部分。但是本书不把相互作用作为重点，因为它比起大脑内的认知过程要容易研究得多，它是外部的和可观察的。为了使认知的研究更加简化，应该注意区别而不是混同文化因素与生物因素。

⑥ 语言与⑦知觉

“我之所以列入了这两个主题仅仅是想提醒一下这两个主题的存在，以免被误解为我把这两个主题搞忘了。但是我不打算在这儿讨论这两个主题，因为由于这两个主题的基本的重要性它们已经受到了足够的重视，实际上可以说是受到了过份的重视。因此在认知科学中时而出现一种倾向，认为认知科学就仅仅是关于这两个主题（在知觉之中特别是视觉）的研究。实际当然不是如此，我认为这种错误想法的流行是很不幸的事。语言和知觉本身是非常复杂的课题，其中混在着认知的许多不同的侧面。认知科学中取出任何一个主题来看都是如此，这些侧面处于相互支持的关系之中，某一个认知领域的研究活动通过其他领域所具有的知识 and 特征变得更加丰富。因此我相信，解释语言和知觉的尝试在

认知科学其他十个主题中取得实质性进步之前是不可能见到成功的。”

⑧ 学习

“学习，是很多人都认为唯此才是关键，而至今仍然是个谜的课题。在心理学的初期和尝试智能人工装置的早期阶段，学习曾经是核心的课题。人们制作了通过相互作用来学习的机器，这可能是由于相信这样作可以最终得到广泛的通用的智能。心理学家所发展起来的人和动物行动的通用理论经常是以“效应法则”或“学习和记忆联合的各种特性”这类基本学习原理为中心构成的。但是这种尝试全都没有结出果实。时至今日，学习的研究不论在心理学领域还是在人工智能领域都已不再被认为是处于中心部分。这是为什么呢？可能是由于为了理解学习，首先必须理解表现的问题，输入（知觉）的问题、输出（实行行为）的问题以及思考和推论的问题。而关于这些问题直到最近我们才开始有了一些粗浅的理解。”

“为了学习我们花费了一生之中非常之多的时间。在某种意义上，不管干什么我们都从中学了点什么。我们之所以至今没能理解学习，就是因为学习不仅仅是知识的积累。知识的积累确实也是学习的一种方法，但还有许多其他必须研究的东西。学习的一个基本形式就是对自己已有的知识进行重构。也就是作为获得关于某种新概念或新知识的结果，关于某种事物的理解在根本上发生重构。此外还有把初步的技能和理解磨炼成流畅高效自如的专家技能和理解的堪称调律的过程。”

“实际上关于学习的研究少得惊人。当然这里所说的是真正的学习，也就是要花几个月，有时是几年的关于复杂主题的学习。在其他地方我曾经推算过可以被称为某种工作的专家的人为了获得从事该项工作所必要的技能，花费的时间不下5千小时。这并不是特别长的时间。一周40小时，一年50周，全用来工作要两年

半。作为成为网球选手、计算机程序员或是语言学家所需要的时间来说还是太少了。在这么长的时间里到底发生了什么？不管发生了什么，肯定是一个极为缓慢的、连续的过程。不管是用药丸子还是讲课的形式，都不存在把知识灌进脑子里的灵丹妙药。在这里有效的仅仅是缓慢的不间断的与主题的接触，同时伴随着经过几次的内心表现的重构、概念的再形成、以及在漫长时间中进行的大量事实积累，仅此而已。”

“学习这东西的相对的重要性已得到充分的认识，而且经常被人谈起。我们知道学习对于孩子们是何等重要，以及儿童向成年人成长的过程是何等的重要。在这个意义上可以说我们都是职业的教育家。所以说到尽管如此，我们关于学习（及其对应物的教育）竟然如此无知，实在令人惊奇。而且在这个场合，这种理解不足来源于想要理解的努力不足。人们常常兴奋地谈论一种梦想的计算机程序。这种程序先把最基本的知识放在里边，而后所有必要的知识都由程序通过动作环境中的实际经验，由自然的学习过程获得。但是关于真要使这种程序成功到底需要多少工作的研究实际上几乎没有进行。很可能由于大家都在暗地里承认这个工作肯定比表面看起来的要远为困难。可能出于对这种令人沮丧的事实的认识，即从婴儿到初步的专家要费 25 年的时间，从基础开始需要 5000 小时的学习和训练。谁会想要一个不先全时运转 25 年就不能用的计算机程序呢？（把 25 年换成 5000 小时也行，谁要是认为能够说服政府的补助金机关或是大学教授会的委员会，可以试试看。当然要加一个条件，即最初几次可能费了这么多时间而结果什么也没学成。）总之，关于学习过程的研究和理解不管过多久也是一事无成，实在令人悲哀。”

学习是不是非要等到表现、知觉、行为都弄清了之后才能开始研究？我看未必尽然，也可能只有在弄清了学习之后，我们才能知道为此需要什么样的表现以及知觉是什么。当我们遇到坚不

可催的碉堡时会想到一句格言，堡垒是最容易从内部攻破的。至于学习的时间问题也不想诺尔曼想象的那么严重，一个二十五年可以学成专家的程序，只须几个月就足以让人明白它不同凡响。我曾经亲耳听到中国一位科技界权威人士声称，如果能作出这种程序，要多少钱给多少钱，只是当时会场上没有一个学者敢要而已。

⑨ 记忆

“关于记忆的心理，表面看好象已经知道了不少，但是还不忙高兴，实际上弄清的问题并不象读者想象的那么多。”

“关于记忆的性质的研究已经发现了几个重要的功能，其中有些是明白的，有些还不大清楚。第一个弄明白的是，人的记忆是人的认知的核心，记忆系统是认知系统的中心部分。但是，不能说我们对于从非常大的记忆容量中进行信息检索时出现的各种难题给与了充分的注意。‘计算机科学’还没有到达面对这种信息检索的真正难题的地点。当使用在获得信息的阶段完全没有预料的形式发问时，为进行适当的回答怎样才能找到必要的信息呢？对于人来说这是极为常见的事，而现在的人工系统却不可能作到。此外，一开始并不知道必要的信息是什么，为什么在检索得到信息时，会知道它就是所要的东西呢？例如，当我们努力回忆很久以前去世的同事的名字，当艾萨克牛顿这个名字出现在脑子里时，尽管要找的名字已经不记得了，为什么我们能断定不是艾萨克牛顿呢？这个例子本身包含了解决问题的关键。但一般的场合并不那么单纯。我们到底是怎样记忆话语、事件、经验之类的呢？或者对现在的议论用更恰当的论点来说，怎样在完全没有预料的场合想起这些内容的呢？”

“关于记忆还有各种各样的不可思议的现象。理解一个词汇的意思所需的时间（例如在读文章时）以 1/10 秒为单位，与此相比，当我们要找到一个在文章中该使用的词汇时，有时要花几小时甚至几天。那么，到底是什么东西使得记忆检索持续几小时甚至

几天的呢?况且与此同时有意识的思考过程早已进展到其他地方,那个词的必要性也早就消失了。现在发生的事件把我们的的心灵带回过去的经验。但是这种带回的方式并不总是理所当然的。按照公认的说法,记忆是相互连合的,而且记忆是按照某种形式形成结构的,例如网络、概念、原型、基础水平、图式、框架、单元、剧本等等。但是为什么是这些结构,它们又是怎样工作的呢?我们有必要理解知识的表现,但这种理解必须包括作用于这些表现的过程。先要弄清象运动记忆、形象、空间信息之类到底是什么?”

“由被记忆的概念之间的联合立刻可以想到的是,记忆结构之间是否有一种类似于导线那样的东西。但是只要稍微想一想,就会明白这种(神经细胞的)导线概念是全然不可行的。这是由于(假如我们假定记忆各自存放在不同的地方)每当我们把新的记忆存放在一个地方时都要从旧的记忆场所跨越一定距离向新的场所蜿蜒伸出几条导线(的生物学对应物)来。要想找到代替导线的东西并不容易。最有希望的候补是考虑给每个场所加上带编号的标签。(不必太在意编号,主要是各个场所都有自己独特的名称就行了。)这样,两个记忆结构之间的连合,只要让它们各自持有对方的编号,然后由某种巧妙的装置利用这个编号把一方与另一方连接起来就行了。这个我们称之为‘地址问题’的问题是在组织某种程度的大规模连合记忆时的基本问题。作为一种克服地址问题的方法, Bobrow 和我提出了按照检索项目的‘陈述’来查访记忆的方案。在发表本书所收集的论文的学会上明斯基提出了别的方案。按照他的方案,连合起来的记忆结构之间实际上连着导线(神经),由这种物理连接而产生的问题是,关于什么和什么之间有可能建立关系需要设定非常严格的制约条件。”

“不过且慢。我们假定了各个记忆存放在各自的场所,但是这个假定是必要的吗?把记忆分布在空间之中是不可能的吗?(请考

虑一下全息摄影)那是可能的。也就是说有必要考虑两种本质不同的记忆结构。一种是我们刚讨论过的分场所的记忆,另一种称为加法型记忆,各个特定的记忆单元相互重叠放在一起,靠各自的图式在必要的时候从重叠的记忆中抽出必要的信息。在这种叠加记忆的范畴中可以包括全息照片、所谓的‘联想记忆’、感知机(及其现代版)。这种记忆方式自动提供了按内容地址方式进行记忆保存和检索的方法,但另一方面又存在着大量本身特有的难题。现在还不能说对于加法记忆已经进行了充分的研究。”

“最后,我们还剩下非常重要的有关记忆系统的功能特性的问题。关于这个问题已经进行了一定程度的研究。例如短期记忆(一次记忆)、作业记忆、记忆的活化等问题。此外还有记忆的各种使用方法的问题。例如记忆的组织化方针、检索方针,还有记忆的排练,也就是暂时记忆的内容通过多次的反复,(在这段期间什么作了些什么?)可能在这个期间对记忆的内容施加了某种操作。暂时记忆到底是一个呢?还是有很多个,还是根本就没这个东西呢?此外,永久性记忆或是作业记忆中作为表现的素材所使用的是形象还是命题呢?先到此打住吧。这种问题还可以举出很多,而且是众所周知的。”

⑩ 行为实行

“运动控制程序在本质上不可能是单纯的,制定这种程序需要与读书、知觉、说话同等程度的认知功能。完成这种程序是很费时间的作业,运动过程越复杂所需时间越长。它还受到同时进行的其他行为的妨碍,需要很长的训练时间。”

“考虑到需要控制的肌肉的数量,再加上基于很多关节和身体的柔性的庞大的自由度,可以认为对所有拮抗的肌肉组进行为使其正确运动的计算是超越可能性限度的作业。”

“人脑之中有非常大的一部分专门用于运动控制。小脑的巧妙机制与运动中枢一样,看起来专注于这项工作。人脑有如此之大

的部分专门用于运动控制，很难想象运动控制可以和高级精神过程的研究无关。感觉系统与运动系统之间有密切的联系，这种密切联系得到神经生理学的支持，心理学上相同的密切联系可能也存在。”

“行为实行的各种问题不是表层的问题，要解决这些问题需要理解相当高度的计算技术问题，而且必须作为与知觉和思考过程在基础层次上具有相互作用的问题来考虑。很可能，我们关于世界的知识的大部分是以应该与世界之间保持什么样的相互作用这种手续型知识的形式存在的，可以想象，知觉—认知—运动图式是统合的记忆结构，硬把它们分离会使它们都被破坏。”

① 技能

“技能？为什么不能认为技能就是和学习、行为实行、记忆之类一样的东西？技能不就仅仅是熟练的行为实行吗？”

“技能，就是学习与行为实行的结合。但是这里还有超越这种结合的东西。这也许可以说是人类认知的本质的一面。现在假如我们想象我们所具有的生物学遗产是以为特殊行动的特殊子系统的形式发展起来的。也可能技能就是这种形式的具有独立的知识源、独立的计算资源、极而言之甚至还有独立的脑子和身体结构的象知识包那样的东西。也可能是这样，也可能不是这样。一如既往，我认为真实在这两者之间的某个地方。我们既不是对所有知识和所有能力作完全相同处理的通用计算装置，也不是相互完全独立的高度专用化子系统的简单集合。实际上，我把各种功能不称为独立的而采用与之对立的可分离的概念。确实我们不能否认在进化过程中作为某种生物学的存在会有功能的特殊化。所以我们是否果真具有可分离的各种技能这件事，可以说是对人类认知的各种理论具有重要影响的重大问题。”

“技能是知识和行为实行的特殊化子系统。熟练者和非熟练者确实用不同的方法来完成课题。这个说法是正确的，但是漏掉了

这种不同的本质是什么这一点。熟练者的所作所为在质和量两方面都与非熟练者不同。Bartlett 在他关于思考的书中说，熟练者与非熟练者主要在时间配合上不同。并且说熟练者有着充裕的时间，他们完成工作时轻而易举、连续平滑、无需特殊努力，在时间上游刃有余。熟练的网球选手在球还没飞到时已经来到了击球位置，熟练的驾驶员象出到飞机前头那样操纵飞机。对熟练者来说，这些困难的工作看起来很简单。与此相比非熟练者总是蹒跚、忙乱、被动。所以对于非熟练者来说，困难的工作确实看起来很难。”

“还有其他的不同。那是注意点的不同。请读者回想起自己刚学开车时的情况。你所受到的教示只是动作和机械的工作原理。这样握住方向盘，脚（离合器）和手（换挡）这样配合操作等等。但是随着你的技术提高，视点也逐步变化。你已经不是向顺时针移动你的手而是在转动方向盘。其后又成为在让汽车转弯，而后又成为让车进入某个路口。最后，当你成为熟练的司机时，你想的只是去银行或去购物。关于自己的所作所为的这种感觉的质的差异是很了不起的东西。当你进入熟练者的领域之后，你已经变得不再意识到你作事时附带的所有操作。看到一个路口，形成进入路口的意图，汽车就忠实地完成你的意图。开汽车成了象走路一样自然的事，汽车就象手和脚那样，成了受你控制的身体的一部分。”

⑫ 思考

“时至今日，认知的研究应该包括思考的研究是不用说也知道的。但问题不在于是否应该包括，而在于应该以什么形式包括。读者可能认为我们对于思考已经知之甚多。我可不那么想。我们对于思考虽然拥有重要的知识，但这些知识主要局限于思考过程中能够用意识把握的部分。而且由于我们对于意识本身还缺乏知识，所以结果我们还不能说我们完全弄清了有意识的思考的作用。”

“应该认真讨论的问题之一是，‘思考’这个题目如果与所有

具体的知识和特定的机制分离而完全独立，作为完全纯粹的抽象活动，在与其他活动分离的前题下，多大程度的研究是可能的。思考的数理理论正带有这种性质，随着人类思考的模型中越来越多地使用数学，看起来把人看成是通用计算装置的倾向也越来越强。”

“但是，如果我们并不是那种通用的存在，事情又会怎样呢？如果我们的思考过程是按照适合于与世界进行相互作用的原则设计的，由经验得到的内心模型是我们主要的推论工具，而我们进行抽象操作所需的时间和在心中保持形式结构的能力略有不足，情况又如何呢？按我的想法，我总觉得人们对于人类的推论机制的可能的形式化特征强调得太过份了，而有忽视立足于非形式化经验的推论模型的毛病。但是当我们接受这种议论时必须十分谨慎。我赞成 Newel 的这种说法：作为认知的很多部分的基础，某种形式的通用的物理的符号处理系统是必要的。我们也许是特殊化的，但另一方面也可能是通用的，可以学习新的能力，可以就新的状况进行推论和规划。尽管如此，为此我们使用的方法也可能仍是本来用来对经验建模的方法，为了计算的方便利用空间布局的各种特性（用形象思维解决非形象问题），而且很可能是利用具有遗传固定线路的特化的空间知识。在这一点上我同意 Johnson-Laird 的意见。”

“环境在思考中扮演重要角色。我们在解某些问题时想象着环境，在解某些问题时利用着环境。密克罗内西亚的航海者们明显地把他们的独木舟的侧杆当作一种模拟计算机来使用，按照星星的位置和独木舟侧方的水的流速的组合，可以在几乎不见陆地的条件下正确航海几百英里。（Hutchins, 1979）我们自己也把铅笔、纸、图、甚至书桌上物品的分布作为思考的外部辅助手段来使用。计算机在某种意义上也是我们认知的人工延伸物，是人类为了扩展人类思维过程而发明的。正如由于有了文字可以免去我们学习

记忆术的必要，有了袖珍计算器学习算术的必要不再那么强烈，有了打字机不必再学钢笔书法那样，一旦到了小型便携计算机人手一台的时候，我们的某些类型的思考也可能不再必要了。（往好里想，我们可以获得相应的沉浸于高水平思考过程的自由。）这里不是讨论这些变化带来的社会影响的场所，只是想指出我们的文化在多么大的程度上依存于与思考过程有关的技术辅助手段。”

关于认知科学与12个主题的关系，诺尔曼说：“我深切地希望认知科学被认为是研究各个有关主题之间复杂的相互作用的学问。而且这种研究必须建立在这样的认识的基础之上：在理解各有关部分之前不可能正确理解这种相互作用，其中任何一个部分都不可能与其他部分完全独立，理解部分需要理解整体，理解整体也需要理解部分。”

在这里我们不厌其烦地几乎全文引用诺尔曼关于12个主题的说明，是为了使读者对认知科学所研究的课题及其困难程度有个比较全面的了解。没有这种背景知识，将很难理解本书的意义。在我所知道的文献之中，这12个主题是对认知科学最全面，最深入的阐述。关于12个主题的阐述回答了本章开头提出的问题：钥匙丢在哪儿了？那儿为什么暗？

第二章 工程认识论

认识论冠之以工程二字,给人的第一个印象就是不伦不类。认识论属于哲学范畴,与工程风马牛不相及。这里所说的工程认识论,不是指认识论在工程界的应用,而是指一种用工程眼光看起来也能满足要求的认识论。这种要求最起码的一点是消除形而上学,这一点也完全违反哲学的习惯,哲学与宗教历来就是形而上学的大本营。所谓形而上学,是指那种凭空立起来,不需要任何论证和支持而当然正确的东西,有点象几何学的公理。形而上学拒绝任何科学的分析、论证和研究。我们所需要的认识论是既作为认知科学的工具,又作为认知科学的哲学层次的表现的认识论。为了满足这种需要我们不得不要求这种认识论中的所有概念都存在着科学的对应概念,也就是不包含与科学无法发生上述关连的形而上学成分。那么我们为什么不把这种认识论称为科学认识论呢?一个原因是,科学这个词由于其现代的声威而被用得比较滥了,毫不科学而自称科学的玩艺儿在街面上比比皆是。所以我们采用比科学条件更为苛刻的工程这个词,其含义是要求认识论的概念不仅是科学的,而且具有工程可实现性。满足科学的要求未必满足工程的要求,例如,如果我们有一种关于认知的算法是指数爆炸的,它仍然满足科学的要求,这种算法可以作为科学的研究对象,但它不是工程可实现的。从这一点上看,工程可实现性比可操作性稍微严格一些,但基本意义一样。

虽然在历史上科学与哲学曾经是浑然一体的,但是现代科学已经与哲学划清了界限,在哲学中可以讨论科学成果,但在科学的领域里混入哲学通常被认为是一种非法越境行为,哲学的结论

不允许作为科学的根据。爱因斯坦说过：“没有认识论的科学，即使可以想象，也是幼稚和混乱的。”任何科学行为，不论是否被意识到，都依存于某种认识论，因为科学的活动是一个认识过程。但是在一般的自然科学领域内部讨论哲学问题是既无必要也不可能的。因为自然科学的研究人员都具有基本上一致的认识论基础，例如相信客观世界是有规律的。已经一致的东西，当然用不着讨论。即使发生争论，也是在科学的领域之外，例如爱因斯坦和波尔的争论，集中于对量子理论的解释和评价，这种争论对于理论本身的形式并无影响。在科学领域里讨论哲学之所以不可能，是由于哲学中充斥着形而上学，而形而上学又是拒绝用科学方法加以分析和研究的东西。在科学领域里探讨不能用科学方法处理的东西显然是自相矛盾的。因此在一部科学著作中专设一章来讨论哲学问题作为科学理论的必要组成部分是非常违反科学常规的。遵循常规的事只管干了就是，作违反常规的事则要承担对其合理性进行说明的义务。

认知科学也是自然科学，但是具有与其他自然科学不同的特点。认知科学与其他自然科学最大的区别在于其研究领域是与认识论完全重合的。二者都是研究人的精神世界和人的认知过程的，只是研究的手法和表现的层次不同而已。由于这种重合，与其他自然科学相比，认知科学与认识论的关系有了质的变化。一般的自然科学与认识的关系是可分离的，作为工具的认识论与作为产品的科学理论可以没有任何相似之处，就象一个机器零件与制造它的工具不需要有什么相似之处一样。反过来，科学哲学的理论体系也可以和它的研究对象即科学的理论体系没有太多的共同点。因此对于研究一般自然科学的人来说，声称自己抱有什么样的认识论观点，或是关于认识论什么也不说，对于他所研究的科学理论都可以没什么影响。但是对于与认知科学有关的学科来说，情况就大不一样了。认知科学和认识论都是对认识的认识，而这

两种认识又是同一个认识。这句话本身就象个绕口令，够令人费解的。我们可以仿照数学里把微分的微分称为二阶微分的方法，把对认识的认识称为二阶认识，简称认识2。而原来的认识叫作认识1。那么上面那句话的意思是说，认识2与认识1是相同的认识。为什么相同呢？因为人只有一个脑袋。我们无法去制造一个超人来实现认识2。甚至这种想法都会导致理论困难，因为认识2如果与认识1不同，那么为了弄清认识2是什么以便制造它的话，我们先得造一个具有认识3的超超人。依此类推，陷入无限退行。需要注意的是，用微分来作比喻是极不恰当的，因为二阶微分是对一阶微分所得到的函数进行微分，而对认识的认识并不是指对前次认识的结果进行再次认识，它相当于对微分符号或微分运算符进行微分，在数学中并不存在这种运算。如果要在数学中找比较合适的对应物，大概只有哥德尔不完全性定理的证明过程。这种理论带有自我陈述的性质，相当于一个“怪圈”。怪圈这个词可能读者并不太陌生，但在我国的报刊杂志上出现这个词汇时通常作为“恶性循环”的代用品，作贬义词用，会给人造成错误的概念。实际上“怪圈”的概念是理解自我陈述系统的重要的核心概念，为了建立正确概念，希望读者阅读霍夫施塔特的著作，例如其中文译本《GEB——一条永恒的金带》（走向未来丛书，四川人民出版社，1984）。

在认知科学中工具与产品是同一个东西，认识论本身就决定了认知科学的成果，反之认知科学的成果也决定认识论。建立这样的理论就象要用一把刀子去削这把刀的柄，不大好办。为此可以作两把一样的刀子，相互削对方的柄，只要削的时候注意使二者保持一致就行了。这就是我们为什么非要在认知科学的书里加上认识论的原因。认识论和认知科学就相当于两把具有某种一致性的刀子。

爱因斯坦在谈到科学家的认识论观点时说：“寻求一个明确体

系的认识论者，一旦他要力求贯彻这样的体系，他就会倾向于按照他的体系的意义来解释科学的思想内容，同时排斥那些不适合于他的体系的东西。然而，科学家对认识论体系的追求却没有可能走得那么远。他感激地接受认识论的概念分析；但是，经验事实给他规定的外部条件，不容许他在构造他的概念世界时过分拘泥于一种认识论体系。因而，从一个有体系的认识论者看来，他必定象一个肆无忌惮的机会主义者；就他力求描述一个独立于知觉作用以外的世界而论，他象一个实在论者；就他把概念和理论看成是人的精神的自由发明（不能从经验所给的东西中逻辑地推导出来）而论，他象一个唯心论者；就他认为他的概念和理论只有在他们对感觉经验之间的关系提供出逻辑表示的限度内才能站得住脚而论，他象一个实证论者；就他认为逻辑简单性的观点是他的研究工作所不可缺少的一个有效工具而论，他甚至还可以象一个柏拉图主义者或者毕达哥拉斯主义者。”（《爱因斯坦文集》第1卷，商务印书馆，1976）。

这段经常被人引用的话到底说明什么呢？有人认为它暴露了爱因斯坦思想庞杂、立场不稳，甚至有人以此指责爱因斯坦本人就是一个哲学上的“机会主义者”。这种看法显然不对。工欲善其事，必先利其器。我们难以想象拿着最坏的工具的人能干出最好的活来。

按照李醒民在《两极张力论》（三原色丛书，陕西科学技术出版社，1988）中的说法，“事实上，爱因斯坦的这些言论，正是他的思想和行动的真实写照。敢于正视矛盾的两极，善于在对立的两极保持必要的张力，恰恰是爱因斯坦的认识论和方法论的一大特征，也是他在科学探索活动中取得成功的秘诀。”按他的说法，爱因斯坦善于在经验论和唯理论、归纳法和演绎法、无神论和宗教感情这些对立的两极之间保持张力。这个说法不乏真知灼见，但是不科学。因为他没有说清楚怎样保持这种张力，只是把它形容

为走钢丝一样的高超技艺，稍一偏离平衡位置，就会发生“翻车”。科学要求重复性，你能干，别人掌握了相同方法也一样能干成的才算科学。如果有人教我游泳，告诉我：“你应该巧妙地使你的手和脚的动作协调起来。”这话不错，但不科学，我听了以后仍然不知道手和脚的动作要领。较为科学的说法是：“根据流体力学的原理，物体在流体中所受的阻力与其速度的平方成正比，因此，为了获得前进的动力，在作对推进有效的动作时要快而猛，作妨碍前进的动作时要舒缓。”这种理论对于改进蛙泳动作是很有效的。从这一点看，李醒民的说法还不如爱因斯坦原来的说法，爱因斯坦毕竟指出了在哪种场合应该使用哪个认识论。

按我的看法，爱因斯坦的这段话至少说明这样一个事实：即使对于一般自然科学或物理学而言，可用来作为工具的完整的认识论体系也还不存在，科学家被迫采用一种技巧随时改变认识论的立场。

从工具的角度来看，当我们进行一项科学研究而缺乏某一种仪器时，可以想到以下的几种方法解决：购（租、借）入、改装、定制、自制。第一个方法最省事，翻翻产品目录，或是到中关村街上去转转，有现成的买一台来就行了。但这种方法的前提是存在这种产品，从爱因斯坦的话可知，满足我们要求的认识论体系还不存在，此法行不通。第二个方法是改装，把不合用的认识论体系修改一下，使其满足要求。比方说我们要求认识论体系不包含形而上学的成份，可以选一个现成的认识论体系，把其中的形而上学部分去掉。这样作就如同拉一头牛来，任你有庖丁解牛的技术，把骨头剔得一干二净，可剩下的只是一堆肉，站不起来，不成其为牛了。以形而上学为框架的哲学，去掉形而上学之后，也就不成体系了。改装也行不通。定制是向专门厂家提出要求，让厂家按特殊要求作一台仪器。这相当于我们找一位专业的哲学家，告诉他我们需要什么样的认识论体系，让他按要求给造一个。但

我估计没有一个职业的哲学家肯接受这个定货单，因为我们的要求是哲学家所不能接受的。就拿不包含形而上学这一条来说，形而上学是哲学家的看家本事，让哲学家放弃形而上学，就如同让舞蹈家捆起手脚跳舞，让木匠作家俱不许用刨子一样。最后剩下的路只有一条，自行制造。当然，不可能从零件造起，只要把认识论新体系的骨架搭好，肉可以从别的体系解下来贴上。我们不是职业的哲学家，充其量算个业余爱好者，但事逼到这一步，只好班门弄斧了。

2.1 东西方文化比较和深层精神结构

从爱因斯坦关于认识论体系的话中还可以作出更深一层的解释，那就是西方的所有认识论体系都存在着类似的片面性。要想建立一个全面的体系，必须理解这种片面性的根源。如果在某种文化基础上建立起的体系具有某种通病的话，我们有理由怀疑这病根就在文化基础之中。而对于挖掘这种病根来说，比较文化学是个有效的工具。如果不能站在不同的文化基础上来看问题，再有批判精神的人也有他的盲点。永远站在一个立场上不移动的话，任何人也看不见他脚底下踩着的那块地方。在某种文化环境下成长起来的人，他的思想体系中的深层精神结构被文化环境所决定。在不遇到异种文化撞击时，这种深层精神结构的存在甚至难以被意识到。即使在异种文化的撞击之中，如果不作用心的分析，通常也只停留在不理解或仅仅看作是习惯不同而已。北京和纽约是姐妹城市，有一次北京市长访问纽约时，纽约的动物保护团体向这位市长抗议北京正在进行的灭狗活动。市长出于礼貌说：我们杀狗是用人道的方法进行的。北京人看了这段新闻可能哑然失笑，我们吃狗肉关你美国人屁事。汉城奥运会时为了避免麻烦，有关当局也让街上的狗肉馆换了招牌，请大家忍几天在开会期间不吃

狗肉。一个微不足道的狗肉问题，其实有其深远的文化背景。西方人认为人和天地万物都是上帝造的，上帝造什么都有他的目的，猫生来是为了吃老鼠，老鼠生来是为了给猫吃，狗生来是给人作伴的，当然不能吃，而且整个世界都是上帝造的，所以这个理儿是世界通用的，因此纽约人管得着北京的狗，这是西方人的逻辑。中国人吃什么是神农尝百草的结果，但神农氏并不是绝对权威，他只是体现了一个原则，吃什么不吃什么只取决于好吃不好吃，爱吃不爱吃，这是一个单纯的实践过程，不需要任何形而上学的理由。中国也有一听说狗肉就想吐的人，但他不会认为自己有权干涉别人吃狗肉。

研究东西方比较文化需要对两种文化有相当深入的理解。西方人研究东方文化的最大障碍是一种西方文化中心主义的思想，用西方人的框架来套东方文化，把东方文化纳入西方思想的某个体系，甚至用研究土人的方法来研究中国人，都不可能了解东方文化的深层。抱着吸收东方文化以改造西方文化的目的而研究东方文化的西方人大概是很罕见的。加之东方文化缺乏明示的体系，很容易被看成是一堆神秘的大杂烩。当然中国人理解中国文化就容易多了，一个中国人可能不知道易经，但至少可以知道按他自己的想法什么可以接受，什么不能理解，可以看到一些在西方天经地义的事拿到中国来行不通。反过来中国人理解西方文化也比较容易。在现代中国西方文化的影响随处可见，楼房大厦、电灯电话、彩电冰箱、西服时装、汽车飞机，不胜枚举，会说英语的中国人比会说中文的英美人不知多多少倍。除了吃饭和深层精神结构之外，我们的城市生活已经彻底的西化了。作为一个科学技术人员，不仅要使用西方文化的产物，就连日常工作中思考问题，也得使用西方的思想工具，不仅看电视，还得知道电视机的原理，以及原理背后的原理，这离了解西方文化深层结构也就不远了。加之西方思想比较明晰，理解起来反倒比中国思想困难

要小。我见到过这样一种看法：古代的思想理论就象一盆黑水，谁往里看都只能看到自己的影子。所以我对于孔子的话的解释是否符合孔子的原意，那就很难说了。到底是孔子的思想弥漫于我周围的文化氛围之中而造就了我的思想，还是我仅仅是借用了孔子的语言来表达自己的想法，这并不是个重要问题。

人在童年时期无批判地甚至是无意识地接受很多价值观，形成其深层精神结构。长大成人之后即使具备批判精神，也只能依据深层精神结构来进行批判，决定接受什么和反对什么，而对于自己的深层精神结构本身则不能损其分毫。对于一种文化来说也是如此。文化的幼年时代即神话时代，以及建立早期文化基础的圣贤时代，是我们研究文化的深层精神结构的最佳参照物。而文化后期的理论，就象成年人一贯的虚伪，往往掩盖了深层的理由。因此进行比较文化学的研究，需要从盘古开天辟地说起。

以基督教文化为代表的西方文化和以儒家思想为代表的东方文化的根本分歧起源于各自不同的创世过程，不同的创世过程导致了不同的世界观与宇宙观。基督教的创世过程是先有上帝，上帝先创造了天地，日月星辰，山川海洋，树木花草，飞禽走兽，然后创造了人类的祖先亚当和夏娃。上帝为他们创造了伊甸园，把这两个无知无识的原人养在伊甸园里，让他们光着屁股过着无忧无虑的生活，并警告他们不可以吃某一棵树上的苹果。但是夏娃受了蛇的诱惑，又引诱亚当吃了那个苹果即智慧之果，从此能够区分善恶，于是上帝作为惩罚把他们逐出了伊甸园，人类从此犯下了原罪。（这个理论颇有令人费解之处，既然伊甸园是上帝造的，上帝为什么在那儿安排苹果和蛇来害人呢？但不这么说又无法解释现在的人世，于是为了符合现实不得不牺牲理论内部的一致性。）但是上帝对人间的事并未从此撒手不管，他安排夏娃和一切女性承担生孩子的痛苦作为引诱男人吃苹果的惩罚，指示人们结婚时离开父母，指点人们迁徙的方向，当世上坏人太多时还要制

造大洪水来洗净人间罪恶，而同时还要给好人诺亚一家提供图纸制造方舟以逃避洪水。他创造不同的语言以破坏人们建造巴比伦通天塔的计划，还要规定人的寿命。他还得让处女受孕以生出他的儿子耶稣基督来教育人民，又得让耶稣代人民受过被钉在十字架上。直到今日他还得不断倾听亿万教徒的祷告和忏悔，等到世界的末日他还得审判古往今来的每一个人决定他们哪个该上天堂哪个该下地狱。作为上帝的工作，这里所列的还仅仅是九牛之一毛。

在中国的神话之中，没有哪一位神明能够象基督教的上帝那样把天地万物自始至终地全部承包下来。中国的神明是专业化的，在完成了他们的历史使命之后或是死了，或是光荣引退，不再干涉后人的事。中国神话中的第一位神明是盘古。在天地未开之时，宇宙是一个名为混沌的蛋，蛋中沉睡的盘古一觉醒来，觉得憋闷不堪，抄起一把斧子使足力气劈开了混沌的蛋壳，于是混沌中的物质自动分离，轻气上升为天，浊气下降为地，天地就此成形。盘古开天辟地之后到哪儿去了，我未见报导，但至少在后来的创造活动之中没有再出现他的名字。人的创造要等女娲来作，关于女娲造人的过程，我听到过两种说法，一个是女娲用泥土捏出了人类，另一个说女娲与伏羲婚配，生育了人类。女娲与伏羲还有别的贡献，女娲炼石补天，伏羲则创造了阴阳八卦，发明了二进制和互补原理。但他们对于自己创造的人类则没有过多的关照。其后的事情都是由人类中涌现出的半人半神的发明家解决的。燧人氏发明了取火的方法，有巢氏发明了房屋，神农氏发明了种植技术，仓颉造字，如此等等。在世界上几乎所有民族的神话中，都有关于大洪水的传说，估计这次大洪水是确实有过，可能是最近一次冰川期结束的结果。中国人没有得到上帝关于制造诺亚方舟的通知，只好靠自己的双手和智慧战胜大洪水。禹的父亲受命治水，他从天界借来息壤，这是一种具有自增殖能力的土，放到什

么地方就能自己增加，可以高效率地筑起堤坝，但是这种堵塞方针仍归于失败，大禹的父亲被杀，大禹顶替了治水任务，他三过家门而不入，带领人民实行了开河疏导的正确方针终于治水成功，因而继承了帝位。

从关于天地创造和人类来源的最古老的神话中，可以比较明显地看到东西方文化最根本的区别就是西方的一神论与东方的多神论的区别，或者说是一元论与多元论的区别。可以说东西方文化深层精神结构的众多分歧皆源于这个区别。

由一神论与多神论而产生的第一个分歧是西方文化的绝对论倾向与东方文化的相对主义倾向。基督教文化中的上帝是一个绝对的存在，就象一个绝对坐标系中的原点，世上的一切都可以由它在这个绝对坐标中的位置，即它与原点的关系来定义。而对于原点本身，则只能承认其绝对的存在，很难进行详细的定义，就如同几何学中的点那样，没有大小，没有形状，由于它本身就是原点，也无法讨论它的位置。上帝可以决定一切，但一切都不能决定上帝，讨论上帝本身的性质是极其困难的，也是得不到解答的。例如我们不能问上帝是谁造的，上帝从何而来，上帝喜欢吃什么，上帝长什么样子，上帝下不下围棋，上帝穿什么衣服，等等这一类的问题。上帝具有绝对的形而上学性，是个不可探讨，不可研究，无法规定其性质的存在。与此相比，中国文化中的各位神明则是相对的存在，对于每一位神明都有他的来历，他的姓名、性别、出生地点，现住何处，他的性格急躁还是温和，是否喜欢听人吹捧，有没有妻儿老小，作什么工作，有何专长，信仰什么宗教，是哪一级干部，等等这些性质都是可以研究和探讨的。

由于上帝的绝对性，西方文化中的人是在以上帝为原点的绝对坐标系中定义的，也就是说人是靠与上帝的关系来定义的，人生只对上帝负责，圣经是道德的绝对基础。在中国文化的多神论的相对体系之中，找不到一个绝对的原点，人不能靠与神的关系

来定义，只能采用人与人之间的关系这种相对的定义，即所谓人伦是指君臣、父子、兄弟、夫妻、朋友这样的人与人之间的关系。所以在中国文化中关于家族关系的词汇异常发达，父、母、叔叔、舅舅、姑姑、姨、爷爷、奶奶、姥爷、姥姥、妯娌、连襟、公公、婆婆、丈母、丈人这类词汇中有很多是其他语言中无法区分的。因此中国的道德观念并不建立于某种神学基础，而纯属人与人之间的关系，所谓君君臣臣父父子子，老吾老以及人之老之类都是基于这种相对关系。中国道德基础中的“仁”字，其意义就是两个人，两个人以上时才发生道德问题，一个人是没有道德可言的。鲁滨逊一个人漂流到孤岛上，还是每个星期天要向上帝祷告，但是如果是一个中国人漂于孤岛，则“仁”和“义”都无对象可施。表面看起来，“仁者爱人”和基督教的“博爱”精神是类似的，但其理论基础完全不同，博爱是建立于上帝爱全人类这样一个形而上学的基础之上，人与人相爱只是上帝的最博大之爱的仿效，一种以形而上学为基础的推论；而仁者之爱，则并不存在这种形而上学的基础，它不是绝对的，并不要求人爱他的敌人。关于这种人与神的关系和人与人的关系的区别，最集中地体现在一种下意识的呼唤之中。当人们遇到令人吃惊的现象或事件时，西方人脱口而出的是“啊！上帝！”，东方人则喊的是“哎呀！我的妈呀！”，妈是与自身关系最为密切的一个人而不是神。

东方文化的相对性，还表现在人、鬼、神之间关系的相对性上。对于基督教文化来说，上帝与人之间的区别是绝对的。上帝创造了人和魔鬼，但人和魔鬼都不能创造上帝；上帝不会变成人或魔鬼，人和魔鬼也不可能变成上帝。与此相反，在中国的神话中神与人并不存在绝对的区别，所谓的神也大都是一些既可解释为神，也可解释为人的存在，伏羲与女娲与其说是神，不如说是最早懂得性行为与生育之间关系的人，燧人氏、有巢氏、神农氏这些名字本身就可以解释为发明了钻木取火的人，发明了房屋的

人，发明了农业技术的人。仓颉虽然被神化为长着三只眼睛，但又是历史上实际存在的人物，是黄帝的史官。一部《封神演义》，即有其周王伐纣的真实历史背景，又说明各路神明的来源，本身就是一个由人到神的过程，是历史上大规模造神运动的记录。这种造神运动是一个范例，但不是某个时代的专利，后世人可以依法炮制，关公也好，妈祖也好，都可以随用随造，直到今日也还有把已故领袖的肖象挂在汽车里保佑交通安全者。如果说造神运动是从人到神的单向过程的话，那么《西游记》则充满了人、神、魔之间的相互转换。天上的神仙及其仆从或坐骑可以被贬或自行思凡下界，既可以居深山大河成为妖魔鬼怪、也可以或生于皇室，或入赘于寻常百姓之家过普通人的生活。普通人甚至猴子、狐狸、老鼠、蛇之类的动物，只要经过修炼，也可以得道成仙，甚至一人得道，鸡犬升天。人与神的相对性不仅体现于相互转化，而且表现于日常关系之中。中国的神界与其说是超越人世的存在，不如说是人类社会的投影，玉皇大帝的统治方式与人间的皇帝并无二致，天国和地狱的官僚机构与中国几千年的体制完全相同。神与神之间的关系仅仅是人与人之间关系的映射而已。因而人与神的关系也不会像基督教那样，上帝在人面前有绝对的权威，而是更为平等的、相对的、类似于人与人之间的关系。西方人如果遇到魔鬼的纠缠，只有划十字或举十字架请上帝帮忙之一法。中国人在遇到妖邪，不仅可以请和尚道士降魔，还可以自己动手，泼上一盆乌鸡狗血或倒上一桶粪尿混合物，都不失为一种驱邪之法，甚至可以见怪不怪，其怪自败。中国人求神，常使用“许愿”的方法，比方说亲人生了病，可以去向神明许个愿，如果三天之内病好了，送给神明一只鸡，如果病没有好转，鸡当然也就不必送了。这种人与神的关系，就象人与人之间那种平等的贸易伙伴关系，许愿就好比签定一个先发货后付款的贸易合同，货到之后付款可以防止受骗亏损，这里对神明的信任是有限的，相对的。中国在送

葬时可以烧纸钱，烧纸作的车船轿马，房屋家俱，彩电冰箱，而西方人则不这样作，因为在西方人看来不论死者是上天堂还是下地狱都是一种绝对的、抽象的存在，至少和人世间没什么共同之处，不需要那些东西，而中国人的天堂可以开宴会，喝醇酒，可以封官许愿，生儿育女，在阴间也有贫富之分，也需要行贿，一切需要当然与人间是一样的。

由于东西方文化中神的绝对性与相对性的区别，中国的神也就不具备基督教的上帝那样的绝对权威，只是一种有事才请的，呼之即来挥之即去的存在，在日常生活中不占据绝对的地位。西方的上帝是渗透于生活全过程之中的，出生时要作洗礼，起教名，吃饭时要祷告，星期天要作礼拜，结婚要由神职人员主持，死前要祷告，葬礼要有神父在场，坟上要有十字架。中国人认为出生、吃饭、结婚都纯属世俗的事情，不需要神明的干涉。中国有句俗语是“宁拆十座庙，不毁一家亲”。这句话说明在中国人心中宗教与世俗之间明显倾向于世俗的价值取向。相同的话在西方的基督教文化背景之中就变得荒唐了，我们不能说“宁拆十座教堂，不毁一家亲”，因为拆了十座教堂会使千百人成不了亲。但在中国文化中，成亲与庙是无关的，拆多少座庙也不会影响人成亲。中国的传统结婚仪式中的拜天地，内容是一拜天地，二拜祖宗（父母），夫妻对拜，其中三分之二是对人的关系，如同“天地君亲师”中有五分之三是对人的。而此处的天地，也不具备上帝那样的超自然的绝对地位，天地仅仅是自然的象征，既有象天理或天意这种人力不可抗的对自然的宿命论，又有象“天机不可泄漏”这种对不可预测事物的认识。中国传统的葬礼之中有时也请和尚道士作些法事，但并不是必不可少的，与此相比更重要的是确认孝子的地位这种人与人之间的关系。在中国人的坟墓上，多数没有任何宗教的标志。

中国文化中神的相对权威由于儒家思想而受到进一步的淡化

和削弱。孔子提出“敬鬼神而远之”，把神与鬼混为一谈，主张可以尊敬而不要接近他们。这里的尊敬是虚的，只是为了避免不必要的冲突和麻烦，疏远则是实实在在的。“子不语怪力乱神”实际上成为中国知识分子的传统思想。在西方，神学可以成为一门严肃的正经学问，在中国，谈论鬼神的书是不登大雅之堂的。在美国，总统就职需要把手按在圣经上宣誓，而在中国的历史上对于笃信某种宗教的统治者一向评价不高。在西方宗教之争有可能上升到战争形态，而在中国历史上找不到一起货真价实的宗教战争，信仰分歧的重要性远达不到成为战争理由的程度。中国古代的农民起义经常利用宗教作为聚众的口实，但一旦起义成功或失败，这种宗教也就完成了历史使命，成为用了就扔的东西。

生存是短暂的，相对的，而死亡则是永恒的、绝对的。由于东西方文化对于相对与绝对的侧重点不同，因而对于生和死的侧重点也有所不同。基督教的许多理论是建立在人对于死亡的恐惧，希望死后灵魂得到拯救这一基础之上的，如果否认人有灵魂存在或认为人在死后其灵魂也随之消灭，那么基督教的很多教诲将失去意义。与此成鲜明对照的是，当有人向孔子问起人死后的世界时，孔子的回答是“不知生，焉知死”。在孔子看来，活着的事还没有完全弄清楚，哪有闲心去管死后的事呢？这体现了一种价值观，即与生活的问题的重要性相比，死后的问题是微不足道的。而在基督教的教义看来，恰恰是死后灵魂的归宿，决定了人应该如何生活。在很多西方学者看来，能够安葬自己的同类，是人与动物的本质性区别。“数千年的过去，出现了一个动人的景象。在科学家看来与野兽具有相同程度思考能力的人带着悲伤埋葬死者。结实有力的手掘出坟墓，用扁平的石头保护死者的头部，给走上死的旅途的人带上动物的肉作食粮，让他带走加工好的火石及其他有着文明征兆的物品。这些无言的信息跨越了不知多少世纪告诉现代人：‘我们是人，也有悲伤，也相信死并不是一切的终结。

在现代人看来我们的面容可怕,但我们也知道人的苦恼和爱情’。”(Louise B. Young *THE UNFINISHED UNIVERSE*, New York 1986)在中国人看来,死亡只是人类所建立的众多概念之一,并没有决定性的特殊地位。中国有句俗语叫“杀鸡给猴看”,说明至少灵长目动物已经有了对于死的恐惧感。在现代中国达到一定教育水准的人并不认为自己将没有坟墓是一个值得认真考虑的问题。

以死为中心的理论必然带有否定人的自身欲望的倾向。在基督教看来人的欲望是受魔鬼诱惑的结果,人生来就带有原罪。“性恶说”是这种理论的必然结果。如果人性不是邪恶的,人就没有必要向自身之外的绝对存在去寻求拯救,那么上帝和宗教的价值就失去了基础。近代弗洛伊德的学说也把人的下意识本能看成是天然反社会的。这种理论在我们看来很不合乎逻辑,如果人的本能是反社会的,那么又怎样才能解释社会是由人创造的这一事实呢?在基督教文化中不会出现这个理论上的矛盾,因为人类社会可以靠人之外的绝对存在,即上帝的理念来加以维持。因而在西方的理论中往往可以看到一种明显的倾向,即千方百计想找出人和动物的本质性区别,而这种区别往往要依存于一种外在的形而上学的力量,人被看成一半是野兽、一半是天使的存在,而人生的意义就在于逐步消灭野兽的成分,增加天使的成分。因此生物的本能受到贬斥,理性的、形而上学的东西受到赞扬。

但是在以人与人之间的关系来定义人的儒家思想体系中,不存在外在的绝对的力量,人的本性必需是善的。中国古代的儿童启蒙教科书《三字经》的第一句就是“人之初,性本善”。从人性善的基础估计出发,就不能对人的欲望采取完全否定的态度,告子的“食色,性也”的说法,就是对人的生物学本能,即“人之所以同于禽兽者”抱肯定的态度。《礼记》所说的“饮食男女,人之大欲存焉”,也是一样的意思。儒家哲学把男女之交视为“人之

大伦”是其值得注意的特点之一。孔子主张“食不厌精”，则是对食欲的充分肯定。饮食不仅作为维持生命所必须，而且也作为一种欲望的满足被肯定，这在一神论的宗教体系中是很难见到的。在中国，孔子的家传菜可以作为餐馆的特色，孔府家酒可以作名酒商标。在西方，我们恐怕找不到一个经营耶稣菜系的饭馆或上帝牌的名酒。这种深层精神结构，可以说是中国饮食文化高度发达的基础。虽然在较晚的程朱理学有“存天理，灭人欲”的说法，但恐怕还是口是心非，朱熹家吃的未必比别人家差。程颐提出“饿死事极小，失节事极大”实际上违反了孟子的教导“取食之重者与礼之轻者而比之，奚翅食重”。从印度传来的佛教基本上否定人的欲望，而中国土生土长的道教则钻研房中之术，提倡性科学。

在东方文化的多神论和相对主义的深层精神结构的基础上，不存在一个绝对的坐标原点，因此真理的位置是无法确定的。真理这个词汇在中文里是一个在西方文化传入之后产生的外来语，在中国古代思想之中找不到其对应概念。中国人关于理的基本态度是“公说公有理，婆说婆有理”。在《西游记》的世界里，没有任何理由认为玉皇大帝的理和西天佛祖的理是一致的。西方合理主义的一般的推论方法是从某个绝对的原理或原则出发，经过演绎过程而确定某个理论是否正确。而中国人常用的推论方法是“没法子”逻辑，即如果你不能提出更好的方法，就只能认为我提出的方法是最好的。表面看起来，演绎的方法更科学一些，而象没法子逻辑这种基于评价的议论方法则由于我们不能保证所有人的评价准则的一致性，显得要困难得多，而且不大可靠。但是在一个不存在绝对准则的体系之中，基于评价的推论是不可避免的，这种推论往往需要依靠某种直觉。例如，“男女授受不亲”是中国古代的一个道德准则，但并不是一个绝对的准则。在《孟子·离娄上》中有这样一段对话：“淳于髡曰：男女授受不亲，礼欤？孟子曰：礼也。曰：嫂溺则援之以乎？曰：嫂溺不援，是豺狼也。

男女授受不亲，礼也；嫂溺援之以手，权也。”这里的关于用评价来破坏原则的方法，是依靠直觉评价，即“豺狼也”这一句。曾经有一位外国人问我：“听说中国人除了自己父亲的肉不吃，什么肉都吃，是这样吗？”我是这样回答的：“在中国的文化中不存在绝对的准则。按照孝的原则，和父亲顶嘴都是不可以的，但是什么都有例外，在刘邦与项羽夺天下时，有一次项羽把刘邦围在城中，并抓住了刘邦的父亲，扬言如果刘邦不开城投降，就把他父亲杀了煮汤喝。刘邦的回答是，你煮好了汤就分我一杯喝吧。”这是一个极端状态下的评价问题，表面看起来，刘邦的行径可以算得上“豺狼也”了，但是如果开城投降，夺天下之大业将成泡影，跟随自己转战南北的众多将士的成功将毁于一旦，相比之下，牺牲“孝”也是不得已的。当然，并不是所有人都按这种价值观行事，例如也有象袁绍这样因为儿子生病而贻误战机的人，但是这种人得不了天下。

演绎推论一般被认为是科学的，而且是工程可实现的，用计算机来证明定理就是利用了演绎的工程实现。但是基于评价的推论是否可以用计算机来进行呢？关于这一点孟子有个很不错的回答。有人问屋庐子食色与礼相比哪个重要，屋庐子认为礼重要，于是那人问：“以礼食则饥而死；不以礼食则得食，必以礼乎？亲迎则不得妻；不亲迎则得妻，必亲迎乎？”屋庐子答不出，去问孟子。孟子的回答是：“金重于羽者，岂谓一钩金与一舆羽之谓哉？取食之重者与礼之轻者而比之，奚翅食重？取色之重者与礼之轻者而比之，奚翅色重？”（《孟子·告子下》）孟子在这里用一个黄金的小首饰和一车羽毛比重量的例子说明了黄金不一定比羽毛重的道理，由此也隐含了一个关于比重的算法。用数学方法表达，可以把这个算法写成如下的式子：

$$y = \text{sgn}(\sum_i w_i x_i + \theta)$$

式中的 w_i 可以看作第 i 种物质的比重，而 x_i 是它的量（体积），由式子可以定出 y ，即应该作（1）还是不应该作（-1）。只要事先给定食、色、礼各自的重要程度 w_1, w_2, w_3 ，再按实际情况定出三者各自的得失 x_1, x_2, x_3 ，就能决定可为和不可为的 y 值。注意 x 可取正值或负值以代表得和失。这个式子实际上是一个最简单的神经元的数学模型，不仅是工程可实现的，而且是一个神经元即可完成的运算。由此可以说明关于评价的直觉并不是神秘的东西。应该注意这个例子只是说明基于评价的推论是工程可实现的。这个式子只是众多评价函数的一种，既不是唯一的，也未必是最佳的。这个式子中的 w_i 我们通常称之为“权”，它与前文中“嫂溺援之以手，权也”或我们通常所说的“权衡利弊”的“权”用同一个字不是偶然的。

基于评价的推论方法可以自动解决明斯基所说的演绎推论中的例外问题，而且不会遇到指数爆炸的困难。但要想建立这种推论体系遇到的一个重大问题则是如何决定“权”的值，即建立评价体系的问题，这将在后文详细讨论。通常进行科学决策时的方法，也是把各方面的局部利益与整体利益、近期利益与远期利益综合权衡之后作出决策。可以想象人或动物的神经系统也可以自动完成这种决策过程。比起演绎来，这种决策更为动物生存所需要。所以一神论的绝对主义虽然重视演绎，但作为一个人，一个动物必然采取权衡利弊的方法进行决策。虽然在美国的基本法律中有明确的无罪推定原则，但也会由于权衡利弊而出现例外，例如对于移民倾向采取的是有罪推定原则。这种基于评价的综合推论方法我想不妨称之为“中庸逻辑”，以区别于一般的演绎逻辑。可以说，中庸逻辑比起演绎逻辑是一种更为深层，更为基本的推论方法，不仅适用于人，而且适用于动物。正由于其基本，往往潜行于下意识过程中不为人所察觉，让人只感到一种直觉式的结论。在日常生活之中，我们说服他人或自己接受某种观点时，也

很少采用从公理和定义出发进行演译的方法，而是多采用“陈说利害”的方法，诉诸人们的中庸逻辑运算能力。

一神论与多神论的分析所引起的另一个根本性的区别是本体论与认识论的区别。一神论由于存在着绝对的原点，所以天地万物都可以作脱离人的认识的只与上帝有关的绝对的定义，在基督教的创世说中，上帝在创造了人之前就已经创造了天地万物，因此即使其后没有创造人类，天地万物的本体论意义上的存在不受任何影响。但是多神论的体系中不存在绝对的原点，不仅人是靠人与人之间的关系来定义，天地万物也只能靠人与物之间的关系即认识论的关系才能定义。对于盘古开天辟地的故事我们用认识论的眼光来看就容易理解得多。盘古开天辟地是人对外观世界进行认识的初始阶段的一个很好的模型。在人认识世界之前，客观世界并未被人为的概念所分割，所以只是一团混沌，或者说是一个未定义的世界，只有当认识的大斧劈开这个混沌，认识主体创造了上下轻重之类的概念之后，天和地才能被定义。中国神话的创世史就是一部认识史。伏羲与女娲生育人类的故事可以说是人类认识性与生育之间关系的开端。这与伏羲作阴阳八卦的说法也是一致的，阴阳是男女的抽象化概念。燧人氏和有巢氏的故事也是人类对于火和房屋的认识过程及创造过程。值得注意的是，神农尝百草的故事不仅是人对农业的认识的起源，其中还包含了人类认识世界的最基本的方法——试错法。试错法不仅是人类认识世界的最基本的方法，也是生物进化的最基本的方法。按照史书的记载，仓颉是黄帝的史官，当时世上造字的人很多，仓颉对字作了收集和标准化的工作，这也是一个认识的过程，即在一个百花齐放的局面之后，必然须要一个标准化过程。这种过程在仓颉之后也进行过多次，例如秦始皇的“书同文，车同轨”后来各次字典的编辑等等。大禹治水的故事也是一个对洪水规律的试错法的认识过程。从东西方的创世记中可以鲜明地看出其本体论和认

识论的不同倾向。

站在本体论的立场上来看，事物是否真的存在是一个至关重要的问题。从对于上帝是否存在这个问题的不同的回答，产生了唯物论与唯心论的对立。但不论是唯物论还是唯心论，都是站在本体论的立场上来回答这个问题的，所不同的仅仅是那个本体论意义上的本原的绝对的存在到底是上帝（绝对精神）还是物质，即绝对坐标的原点放在何处的问题。但是孔子对于神是否存在这个问题的回答却是“祭神如神在，不祭神如神不在”。这个答案如果站在本体论的立场上是难以理解的，觉得它根本不能构成一个答案，人们会问“那么神到底存在不存在呢？”。孔子的关于神的存在说法只有站在认识论的立场上才能理解。孔子所说的神的存在，是一种认识论意义上的存在。也就是说，神是人在认识过程中为了满足感情上的某种需要而创造的一个概念，它也象其他概念一样只存在于人的关于客观世界的内心模型之中，当人们为了某种感情上的需要而对神举行某种仪式时，他必须假定神是存在的，如果不承认这个假定，祭神就成了无意义的举动，但是在不祭神的时候，这个假定并不重要。从认识论的立场来看，本体论意义上的绝对的存在与否并不是一个重要的问题。对于习惯于站在本体论的立场上来看问题的人来说，这种看法是难以理解的，后文将进一步详细讨论这个问题。

站在认识论的立场上来看，在本体论意义上的有神论与无神论的区别并不是一个太重要的问题。只要你开始议论神是否存在，每一个参加议论的人在头脑中都已经有了关于神的概念，否则议论就无从开始，但从认识论的立场上看，这已经构成了神作为一种概念，作为认识论意义上的存在的证明。但是在一个以多神论为基础的相对主义的思想体系之中，神不具有绝对坐标原点那样的重要作用。因此对于鬼神之类的问题进行讨论不会带来任何重要的成果。正是基于这样一种认识，孔子主张“敬鬼神而远之”。这

里的“敬”只是一种虚伪的态度，因为如果明确地表示不敬，则很可能和相信鬼神的人（这种人在当时是很多的）发生冲突和争论，而这种争论又是注定没有什么意义的；而“远”则是孔子所真心渴望的，对于无意义的问题，理当退避三舍。因此“子不语怪力乱神”自古至今都是儒家文化圈知识分子的共同行为准则。在中国，宗教与迷信之间并无明显的界限，宗教通常只在教育水准较低的人之间流传。从孔子对鬼神的态度可以看出东西方哲学的一个深层的区别，西方哲学中的认识论是建筑在本体论基础之上的认识论，而东方哲学中的本体论，是建筑在认识论基础上的本体论。不理解这一点，很多现象是无法解释的。

2.2 科学与形而上学

中国古代有个故事叫“杞人忧天”。杞国有个人整日里担心天会塌下来，大地会垮掉，弄得自己没地方呆了，结果患了忧郁症，吃不下饭，睡不着觉。聪明人听说了这件事就去开导他：天是不会塌下来的，天只是一些很轻的气体，即使掉下来也砸不着人。那么日月星辰掉下来怎么办？日月星辰也只是一些会发光的气体，也是掉不下来的。那地要是垮掉怎么办呢？地是由土和石头堆满了的，没有地方可垮。经过这样的讨论，杞人总算放心了，用心理咨询治好了忧郁症。“杞人忧天”这句成语通常是指为没有必要担忧的事而担忧。

因为害怕天崩地裂而患忧郁症令人感到可笑，但是探讨我们居住的大地的稳定性决不是无意义的事情。大地是由谁来支撑着的呢？关于这个问题，自古以来有很多人考虑过，提出了各种各样的答案。这些答案中很大一部分是一种叠层结构的模型。在这种模型中，大地是由某种巨大无比的生物驮着的，这个生物可以是大象、鲟鱼、鲸、乌龟等等，有些则不只是一种生物，其中一

种又由另一种支撑着。但是这种叠层结构的“大鱼模型”有一个明显的理论上的弱点，那就是无限退行。如果我们相信大地是由大鱼驮着的，那么又可以问大鱼又呆在哪儿呢？大鱼呆在海里。那么海水又盛在哪里呢？海水盛在大碗里。大碗又放在哪儿呢？……如此等等，引起一个永远回答不完的无限的问题系列，为了避免这种困境，需要在某个地方把问题打住，也就说到某个地方不允许再问下去了，比方说碗放在哪儿是不许问的，不仅不能讨论，而且也不需讨论，我们必需不问理由地相信大碗是牢不可破而且永远不会打翻的那种绝对的存在，这就是形而上学的必要性。为了使 ourselves 不陷入杞人忧天的状态，我们不得不接受某种形而上学的信仰，例如大多数的科学家都相信客观世界是有规律的，而这一点是用任何方法都不能合乎逻辑地加以证明的。

“大鱼模型”不仅是一种宇宙模型，而且是一种建立理论体系时所采用的结构，一种如图 2.1 所示的直线式的理论结构。图中的

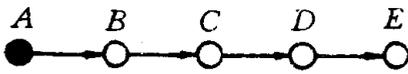


图 2.1

的 A 代表大碗，B 代表海洋，C 代表大鱼，D 代表大地，E 代表大地上的其他东西。这种立足于大鱼模型的结构的最重要的特点是明确

的方向性。例如在大鱼模型中有绝对的上下关系，只能是大碗盛着海水，不能是海水盛着大碗，只能是大鱼浮在海面上，不能是海水浮在大鱼上，上下一颠倒，整个体系就崩溃了。图 2.1 中用箭头来表示这种支持关系，即只能是 A 支持 B，B 支持 C，而不能反过来。在直线结构的理论体系之中肯定存在着象图中的 A 这样的起点，它不受任何支持，是一种形而上学的存在。整个理论体系的稳固性依存于我们对于起点的形而上学的信仰，这种信仰越坚固，则理论体系越可靠。在一神论的文化背景中，这种形而上学的起点即绝对的原点是很容易找到根据的，上帝就是绝对原

点存在的证明。

在科学体系中经常使用的欧几里得范式就是这种直线式理论体系的典型例子。图 2.2 是欧氏几何学的简单图式，图中的黑圆点代表公理，黑方块代表基本定义，圆圈代表定理。有箭头的线代表证明的支持。在欧氏几何的体系之中也存在着绝对的方向性，即只能用公理、定义和前面的定理来证明后面的定理，而不能后面的定理证明前面的定理。这个严格的规定保证了理论中不存在循环论证，但也注定了公理和基本定义的形式而上学性。“直线式”理论

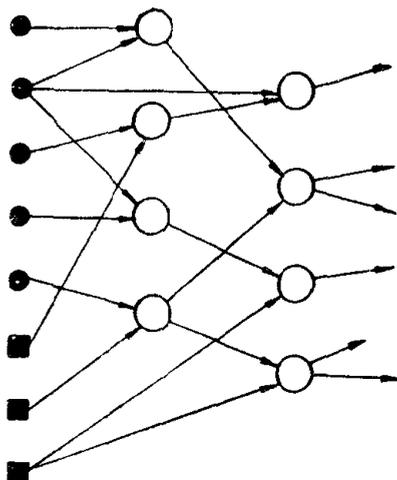


图 2.2

体系是一种简化的说法，用比较严密的说法应称为有方向的理论体系。理论体系中各定理之间的证明关系是一个图论意义上的有向图，它必须满足一个条件，即从任一个节点出发，沿着顺方向前进，都不可能回到原来的节点。我们称满足这个条件的有向图为无环有向图。按欧几里得范式所建立的理论体系的演绎结构都是无环有向图。可以证明，在有限的无环有向图中，从任意的节点出发，沿箭头的逆方向上溯，一定能到达起点，即在一个不含循环论证的理论体系之中，一定存在着不能在这个体系之中被证明的成分，即形而上学的成分。牛顿力学中的绝对时空观，上帝的第一推动力，力的定义等就属于这种成分。

在多神论的文化背景之中，不存在绝对的原点，因此往往不能形成一个大规模的象欧几里得范式所代表的那种理论体系。理论往往各自孤立分散，缺乏严密的内部关系。以中国古代的数学

成果为例，虽然在具体问题的解法或算法上可以属于当时世界的领先地位，但是从系统性、体系化的角度来看，从未出现过象欧氏几何那样的有严密结构的大规模理论体系。其他科学技术领域也可以看到相同的现象。如果要找一个“大鱼模型”那样的比喻来说明多神论背景下众多理论之间关系的话，可以把这种关系比喻成“小行星模型”。众多的理论就象宇宙中漂浮着的小行星，相互之间没有明显的支持关系或一致性，理论之间即使存在着逻辑上的矛盾，也不会导致这种“体系”的崩溃，因为这种“小行星模型”并不意味着相互无矛盾的理论体系。在多神论的文化背景之中，我们没有理由认为各路神明的想法会是完全一致的或相互无矛盾的。这也就是现代科学为什么只能产生于一神论的文化背景之中，而不可能产生于多神论的文化背景的原因。在多神论的文化背景中，人们缺乏对大一统的理论体系的追求，容易容忍理论之间的不一致性。曾经有一位西方人在文化革命中来到中国，敏感地发现电台里早晨放的歌曲里说：“他是人民大救星”而晚上放的歌曲里说“从来就没有救世主”。当然，中国人也并不是没有人发现这个显而易见的矛盾，我认识的一位数学教师就为说了这个矛盾而在文化革命中挨了整，但是一般来说中国人并不把这个问题看作一个逻辑学问题，而是看成一个感情问题，不是用演绎逻辑而是用中庸逻辑来处理。

达尔文提出进化论之后，在一次关于进化论的演讲会上，一位大家闺秀听了演讲吓得晕过去了。这种现象用当时欧洲女子由于束腰和缺乏运动而体质较差来解释是不充分的。在一神论的大鱼模型的理论体系之中，上帝造人是整个社会道德伦理体系的基础，相当于支持大地的大鱼。否定了这一点，就象大鱼突然动起来，整个大地的稳定性都要受到威胁，相当于支撑社会的支柱突然倾倒，这不是杞人忧天，而是非常现实的恐怖。在这种强烈的精神冲击之下，吓晕过去是不奇怪的。但是进化论传到中国时，甚

至听不到反对的声音，在众多的小行星中加上一颗，并不是什么大不了的事。知道人是猴子变的，并不意味着明天就要生活在树上摘果子吃，在中国即使是严重的心脏病患者，听听进化论也是对身体无害的。而在大鱼模型的理论体系之中，任何危及大鱼的理论都是骇人听闻的。布鲁诺和哥白尼的遭遇是这种理论体系的必然结果。

现代科学起源于一神论，所采用的基本范式是欧几里得范式。在这种直线式的演绎体系中，必然存在着形而上学的起点。然而，形而上学却又是与科学水火不相容的东西。科学的基本方法是论证和验证，但形而上学则恰恰是拒绝任何论证与验证的，它只要求无条件的信仰。因此，有形而上学的地方就没有科学，形而上学处于与科学天生对立的位置，在科学发展到一定深度时，拒绝形而上学的口号便不只一次地被提出来了。早期的科学，如同牛顿力学中上帝的第一推动力那样，还需要一根形而上学的脐带与神学相连接，但随着科学自身的发展，逐步形成了自身的文化背景，这根形而上学的脐带也就逐步萎缩、断裂了。

科学排斥形而上学的方法之一是淡化理论起点的形而上学性。欧氏几何学的起点称为公理，意思是不经证明而可以认为是正确的理由。而在现代数学中，通常使用基本假定来代替公理这个词来代表理论的起点，它不再具有形而上学地保证正确性的功能。基本假定往往表示为了正确使用这一理论，问题的环境所应满足的条件。在物理、化学这类与现实世界关系更为密切的理论中，理论的起点有时被看成一种约定，有时仅仅标志着学科领域的边界，某一学科的理论起点常常可以在其他学科中论证和验证。淡化理论起点的形而上学性，似乎使得理论的可靠性降低了。如果认为理论的起点仅仅是假定或约定，似乎会造成过大的自由度，使得理论带上了过多的任意性色彩，那么对于理论的约束又从何而来呢？在我们对一种理论的看法之中，通常隐含着反方向

的、也就是自后向前的论证，或者说是一种工具主义的态度。例如我们接受或承认欧氏几何学，并不仅仅取决于我们对作为理论起点的几个公理的形而上学的信仰，还包括很多其他理由，包括理论体系结构上的美感，理论体系中众多结论的有用性，理论体系与其他理论体系之间的无矛盾性等等。也就是说用整个理论体系的成果去评价它的起点的合理性。这就象我们盖一个楼房之前未必知道地基的稳固程度，但在这片基础上盖起了很高的楼而且多年仍不倒塌，我们就有根据说这个基础是足够坚固的。当我们这样考虑问题的时候，实际上是暗中犯了一个错误，就是循环论证，我们在用前题证明结论之后，又企图用结论去证明前题。但是要想从科学的理论体系之中彻底清除形而上学，循环论证又是不可避免的，这里我们需要仔细加以区别，什么样的循环论证是可以允许的，什么样的循环论证是不能允许的。

我们把循环论证分为两种，一种叫作小循环论证，一种叫作大循环论证。小循环论证是不能允许的，而大循环论证则是不可避免的。举个例子来说，有 A 和 B 两个人， A 证明 B 没有撒谎， B 证明 A 说的是真话，这就构成了一个典型的小循环论证，很有可能 A 和 B 都在撒谎。但是如果一个城市或是一个国家所有的人都认为上帝存在的话，虽然让一个无神论者来看，他们全在撒谎，但是当这个无神论者进入这个国家，就会被所有人认为是个撒谎者。小的谎言容易被揭穿，而弥天大谎是非常坚固的。

如果把现代科学的众多学科放在一起看成一个总的大理论体系，那么在这个体系之中大的循环论证可以说比比皆是，为了解释物理学中的浸润现象我们要用化学键的理论，而为了解释化学键理论又要使用物理学中关于电子的理论，我们无法分清到底是物理学支持化学还是化学支持物理学，它们之间是相互支持的，而相互支持的关系也就构成大的循环论证。我们按照某种约定和假定通过某种理论的演绎过程，预言某种现象，而当这一现象被实

验证实，我们又由此推测原来的约定和假定是正确的，这就与演绎过程构成循环论证。在科学哲学的议论之中，常常使用数学工具，例如概率论，但科学哲学的目标是论证科学活动的合理性，而数学又是科学的一个部分，把科学与科学哲学放在一个体系之中来看，就可以发现很多大的循环论证。

与一神论的“大鱼模型”和多神论的“小行星模型”相比较，排斥了形而上学的现代科学相当于“地球模型”。与“大鱼模型”不同的是，“地球模型”不需要也不存在形而上学的支撑，只存在局部的，相对的上下概念，不存在全体的，绝对的上下概念，地球的各个地点，上和下所指的方向是不同的。我们不能把自己所处位置的上下方向绝对化，而认为地球背面的人是头朝下站着的。组成科学地球的各个部分不存在绝对的起点和上下支持的关系，各理论之间相互支持，相互一致，相互无矛盾。因此，作为一个科学理论的必要条件之一就是在这个理论要与现有的科学理论有密切的相互支持关系而不相互矛盾。燃素学说、活力论、心脑二元论这些理论之所以被排斥于科学领域之外就是由于不满足这一条件。燃素学说是燃烧现象的一种解释，认为可燃物中含有一种叫作燃素的物质，在燃烧时释放出燃素。但是有些物质在燃烧之后重量会增加，这就导致某些燃素具有负质量的推论，这与物理学的已知理论是矛盾的。活力论则认为生物之中具有一种无生物所不含有的“活力”或“生命力”，这种理论的致命缺陷就在于，所假定的活力是无生物所不具备的，那么现有的一切关于无生物的物理的、化学的方法对于测定或探索这种活力都是无效的，这就隔绝了关于生命的研究与现有科学理论体系的相互支持的关系，使得关于生命的理论变成孤立于科学地球的小行星。心脑二元论则认为精神世界由于具有“能动的”性质，是不能用物质的任何一种运动形式来加以说明的，心灵是独立于大脑的另一种存在，关于人脑的信息处理理论不可以解释精神现象。心脑二元论

不仅在说明精神与脑的相互作用上陷入理论内部的困难，而且不得不承认“机器中的幽灵”是不可知的⁽⁶⁾。这种理论当然只可能引起某些哲学家的兴趣，而不能成为科学研究的对象，其致命的缺陷与活力论是一样的。

现代科学的“地球模型”与多神论的“小行星模型”相比，其优点就在于地球比小行星大得多。虽然地球和小行星都漂荡在宇宙空间之中，但生活在地球上显然比生活在小行星上稳当得多，小行星的引力太小，稍有不慎就会掉下去。地球的引力大，随便你怎么翻跟头打把式也不会掉下去。一个孤立的小理论很容易被破坏、被遗忘或被其他理论所取代，但是象现代科学这样庞大而内部一致、各理论之间密切关连的理论体系，则是极为稳固的，任何人没有能力另外创造这样一个大体系来取而代之。与小行星相比，地球上的资源丰富得多。孤立的小理论很难发展，而与现代科学相互一致、密切相关的理论则可以发展很快、一方面可以得到原有理论的支持，一方面受到其他理论需求的刺激。象微积分与牛顿力学，黎曼几何与广义相对论，光学及电子光学与生物学，化学与分子生物学等等，无不呈现这种相互促进的关系。

以地球模型作为形象的现代科学体系不再是一个无环有向图，而是一个无向图，任何一个支持都是双方向对等的。一篇科学论文的评价，其被引用次数是一个重要参数。这意味着一个科学理论能够越多地支持其他理论，这个理论本身也就越多地被支持。在一个无向图中不存在起点，因此形而上学是不必要的。

由于科学本身与形而上学对立的性质，加之逆向论证的可能性，在科学领域中存在着自发的排斥形而上学的倾向和能力，科学家自然而然地采取看起来象个机会主义者的行为方式。作为一个科学家，不管他信奉哪一种宗教或者主义，他都不能在科学活动中把这种形而上学的东西拿来作为理论的依据。作为某种理论的开端，科学家有时也难以避免某种形而上学的口号，但这些口

号仅仅起到起点或说明哲学背景的作用，对科学理论决定优劣取舍的，并不是这些口号之争，最终是要看在这些口号下取得什么成果，而且一旦成果被多数人接受，口号自身也就失去了意义。例如在认知科学的领域中就存在着符号主义、连接主义、计算主义、神经达尔文主义等等。众多主义的存在，只是说明认知科学还没有找到一个能够被公认的核心理论，主义只是争论尚无结果的标志，一旦公认的理论建立起来，主义就没有必要了。所以在科学领域中没有牛顿主义，没有门捷列夫主义。这些理论已经被接受，有了充分的科学支持，不再需要主义了。随着科学的发展，各种主义经历着自生自灭的过程。

科学对于形而上学的免疫力扎根于科学自身的价值观体系。形而上学通常代表某种绝对化了的价值观体系。在科学领域之外，例如在日常生活之中，避免形而上学常常是很困难的，这主要是由于人们通常不掌握对于价值观体系进行分析和评价的方法，价值判断有时在潜意识中完成，人们甚至意识不到。在一个均质而变化缓慢的社会之中，各人之间的价值观体系差异较小，冲突不很明显，但是在一个多种宗教共存而关连日益加深，并由于科学技术日益加快的发展使得社会结构也随之剧变的现代世界上，基本价值观体系的不同所引起的宗教冲突，种族冲突、代沟现象等会变得激化起来，使得关于价值观的科学体系的探讨变得越来越难以回避。

2.3 从真理到善理

“科学的目的在于追求真理”，“人生的目标是追求真、善、美”。这样的说法古往今来都是得到大多数人拥护的，但是真理这个概念作为最大的形而上学总是把认识论、科学哲学和其他与真理有关的理论引入困境。谁也无法否认追求真理这个口号对现代

科学发展的巨大推动作用，但是在科学昌明的今日，也没有任何一个科学的理论被人们公认是真理，甚至没有一个科学家宣称自己创造的理论是真理。真理这个概念的魅力以及真理的概念所造成的理论困难都起源于真理的绝对性即形而上学性。真理是绝对的，因此具有唯一性、普适性和永恒性等优越的性质。真理只有一个，所以我们只要掌握了真理，那么可以简单地宣布一切与真理不同的理论都是谬误。真理放之四海而皆准，不随空间位置而变化，免去了在不同地点选用不同理论的麻烦。真理是永远正确的，一旦掌握了真理，就可以一劳永逸的用下去，永远立于不败之地，对未来的一切了如指掌。这个理想的目标对于所有人来说，无疑具有极大的吸引力。即使我们还没有找到真理，或者哪怕我们永远也不能得到真理，只要真理是客观存在的，那么它就能够成为大鱼模型理论体系的可靠的支撑点，足以令人安心，不必杞人忧天。

但是，一切绝对的理论都逃不脱自相矛盾和陷入悖论的命运。如果我们相信上帝是全能的，那么就会面临这样一个难题：上帝能不能造一块大到他自己也举不起来的石头？民主的绝对化会导致民主悖论，即由民主程序合法选出的独裁政权是否合乎民主的原则？自由的绝对化会导致自由悖论，即我们有没有妨碍他人自由的自由？如果把生命的尊严绝对化了，那么吃一颗菜、一粒米都是剥夺了一个生命，这样人的生命又怎么维持呢，以科学的生命观来看，一个什么肉都吃的人和一个待斋把素的人只是五十步和百步的区别。正由于牛顿的绝对时空观在火车思考试验中出现了矛盾才导致了爱因斯坦的狭义相对论。哥德尔的不完全性定理对于建立完备而无矛盾的数学体系的梦想——希尔伯特纲领来说是致命的一击。

真理这一概念依存于一神论的哲学背景，是大鱼模型的深层精神结构所必要的。如果我们认为整个天地万物都是上帝造的，而

且上帝是有理性的，那么上帝在创造世界时以某种理性的原则作为宇宙的蓝图就是很自然的。所谓客观存在着的真理，正是这张“上帝的蓝图”。而人也是上帝创造的，上帝参考自身创造了人，也可能给人装入了一部分“上帝的蓝图”，因此人能够认识世界就不是不可思议的了。但是人所掌握的不是完整的蓝图，上帝留了一手，人的认识就是凭着残缺的蓝图向着上帝的蓝图进军，这也就是追求真理的过程，是本体论基础上的认识论。上帝的蓝图有各种不同的版本，柏拉图的理想国、黑格尔的绝对精神、诸如此类。而康德的先验范畴就相当于不完全的蓝图。如果不接受上述的前题和推论，我们就很容易产生和爱因斯坦相似的困惑：“我觉得最难以理解的事就是这个世界竟然是可以理解的。”

在进一步探讨真理概念所引起的困难之前，在这里要插入一点说明，即为什么在本书中选择卡尔·波普尔(K. Popper)的哲学作为西方哲学中主要的议论对象，而如果把波普尔的哲学思想作为西方哲学的主流或代表显然是不会得到公认的。选择波普尔有几个原因。首先是因为波普尔所研究的科学哲学与本书所探讨的认识论有密切关系，可以说本书受波普尔哲学的影响是比较多的，在前文中已经不加说明的引用了波普尔的一些看法，例如“机器中的幽灵”，“民主悖论”等等。其次是由于波普尔的哲学理性的成分较多，与科学的关系密切，较容易对其进行理性的探讨。第三个原因是波普尔虽然对整个西方哲学体系进行了批判，但它本身又是一个西方哲学，因此本书中对于波普尔部分观点的批评对于很多其他的西方哲学也是适用的。在这一点上，波普尔的哲学思想可以作为西方哲学的代表。下面我们引用一些文章来看看在科学的真理性(也就是认识的真理性)方面所面临的问题，以及他是怎样解决这些问题的。⁽⁷⁾

“长期以来，至少是自牛顿以来，对自然规律的研究一直被当作科学的主要任务。但是科学家们沿用的方法首先是由弗兰西

斯·培根加以系统地描述的，尽管他的说明自从他那个时代以来受到了很大的限制、增补、精练化和复杂化，但是从十七世纪到二十世纪，由他所开创的传统中的某些东西几乎被所有具有科学头脑的人所接受。他的方法是这样的：科学家从实验开始，其目的是为了对介乎我们有知和无知之间的领域中的某一点的现象进行仔细的控制和细致的比较，有系统地记录所发现的东西，也许还要公布出来，经过一定的时期，他和此领域的其他工作人员收集到了许多相同的、可靠的材料，随着这种材料的增多，事物普遍的特征开始显示出来，个人开始形成普遍的假说——关于规律特征的陈述，这些陈述符合所有已知事实，并解释这些事实如何具有因果上的相互联系。单个科学家试图寻找例证来进一步证实他的假说，如果他成功地证实了它，那么他就已经发现了另一个科学规律，这个规律将会揭示更多的自然界奥秘。于是新的突破口出现了——也就是说，新的发现适用于被认为能够产生新知识的领域。因而增加了大量令人激动的科学知识，我们无知的领域退缩了。求知过程又在新的领域重新开始。”

“以收集特殊例子的观察现象为基础而建立普遍陈述的方法通常叫做归纳，并且被看作科学的标志。换句话说，归纳方法的使用被看作是科学与非科学之间的划界标准。基于观察和实验例子——简言之，基于事实——的科学陈述与其他任何种类的陈述（无论是基于权威还是基于情感、传统、思辩、偏见、习惯以及其他根据）形成了对照，只有它才能提供可信和确然的知识。科学是这种知识的身躯，科学的增长在于把新的确然性增加到现有确然性的总体之上，科学的增长便是在这一无限的过程之中实现的。”

“对这种方法，休谟提出了一些疑问。他指出，单个的观察陈述不管数量多大，他们在逻辑上不可能蕴含无限制性的普遍陈述。假如我观察到在某种场合事件 B 伴随事件 A ，在逻辑上并不意味着在任何另一场合事件 B 也将伴随事件 A 。从两个这样的观察中

也不能得出这样的结论，从 20 个、2000 个这样的观察中也不能得出这样的结论。休谟说，如果这种情况经常出现，我也许会期待下一次某 B 将伴随某 A，但这是一个心理学事实，而不是逻辑学事实。太阳也许在我们所知道的过去每一天的第二天出现过，但这并不意味着明天它必将会出现。假如有人说：‘噢，是的。但是从确定的物理学规律出发，如果这些规律所应用的条件和现在的条件相同，那么我们就可以预告太阳在明天升起的准确时间。’我们至少有两点可以将他驳倒。首先，我们已发现的物理规律在过去起作用这一事实在逻辑上并不意味着它们在将来继续起作用。其次，物理学规律本身是普遍陈述，但是被用来证明普遍陈述的已观察到的事例在逻辑上并不包含普遍陈述。因此，这种为归纳辩解的企图乃是通过把归纳的有效性看作自然而然的东西，用未经证明的假说进行辩论。全部科学都假定自然界的规律，假定将来在自然规律被认为起作用的任何方面和过去相同。然而没有一种方法能够保证这种假定。它不能由观察来确定，因为我们不能观察未来的事件。而且它也不能由逻辑证明来确定，因为从一切过去的将来与过去的过去相同这一事实出发，不能得出一切将来的将来与将来的过去相同这一结论。休谟本人得出的结论是：“尽管没有方法证明归纳程序的有效性，但是，由于心理上的原因，我们会情不自禁地产生这方面的设想。既然它们在实践中似乎行之有效，我们就要按照它们行动。然而，这意味着科学规律没有理性的可靠根据——既无逻辑根据，又无经验根据，因为每一个无限制的普遍性科学规律都在二者的彼岸。”

“归纳问题又叫作‘休谟问题’，从他那个时代到现在，这个问题一直困惑着哲学家们。C. D. 布罗德 (C. D. Broad) 称它为哲学的家丑。伯特兰·罗素在他的《西方哲学史》一书中写道：‘休谟证明了纯粹的经验主义不是科学的充实基础。但是，只要承认这一原理，其他一切都能按照我们的全部知识基于经验这个理论

往下进行。必须承认，这严重违反纯粹经验主义，非经验主义者的人或许会问，如果一种违反是许可的，为什么旁的违反就得禁止。不过，这些都是由休谟的议论间接引起的问题。他的议论所证明的是——我认为这证明是无法辩驳的——归纳是一个独立的逻辑原理，是从经验或其他逻辑原理都推论不出来的，没有这个原理，便不会有科学。’”

“整个科学和整个事物必须建立在根据之上，而根据的有效性是无法证明的，这使得人们窘迫不堪。它使得许多经验哲学家成为怀疑论者、非理性主义者和神秘主义者，特别是某些人因此走向了宗教。实际上所有的人都想说：‘事实上，我们不得不承认，严格地说，科学规律是不可证明的，因而不是确然的。即便如此，它们的或然度通过每一个证明例子而提高了；除了整个已知的过去，世界延续性的每时每刻都带来了无数的、成千上万的证明事例——从没有一个相反的例子。因此，即使不是确然的，他们在能够设想到的最高程度上也是可能的；如果不是在理论而是在实践上，那么这与确然性的差异就不容易察觉出来。’几乎所有的科学家在对其所作所为的逻辑根据进行思考时都遵循着这一态度。对他们来说，极为重要的是：科学带来好处——它管用，产生一系列无穷的有用结果，他们不再去纠缠那显然无法解决的逻辑问题，而以从事更多的科学研究、取得更多的结果为乐。然而，在哲学上对这些问题进行更多的思考时，就遇到了很大的困难。对科学家和一般的哲学家来说，归纳对人类知识的根据本身提出了一个尚未加以解决的问题。在这一问题得到解决之前，整个科学无论其具有怎样的内在一致性还是外在有用性，都必然以某种方式游移于半空之中，找不到固定的大陆。”

“波普尔创造的成就为归纳问题提供了一个可以接受的解决方法。在这一过程中，他否定科学方法整个传统观点（本章已对这一传统观点勾画了一个轮廓），代之以另一种方法。如果可以设

想这个成就是根本性的成就，那么它具有的意义超出了原有的范围，导致了对其它一些重要问题的解决。”

“波普尔的解决方法首先指出了证实与证伪之间的逻辑不对称现象。用陈述的逻辑可以表述为：尽管报道白天鹅观察事实的观察陈述不能使我们在逻辑上得出普遍陈述‘所有的天鹅是白的’，但是报道黑天鹅的唯一报道事实的唯一观察陈述能够使我们在逻辑上得出‘并非所有天鹅都是白的’这一陈述。在这种重要的逻辑意义上说，经验概括尽管不能证实，但可以证伪。这意味着科学规律尽管不能证明，但可以检验：通过反驳它们的一系列尝试，它们能够得以检验。”

“现在我们可以理解为什么波普尔一贯坚持认为我们称之为知识的东西，本质上永远是暂时的。任何时候我们都不能说我们现在‘知道的’东西是真的，它总有可能被证明为错误。实际上，一时一刻获得的知识最终又被证明不是这样。这是人类思想史上的一个基本事实。因此，如果仿效科学家和哲学家，走他们一直走过的老路，即力图证明一种理论为真理或为我们所信奉的某种理论进行辩护，就大错而特错了，因为这种尝试在逻辑上是不可能的。然而，可能最重要的是，我们所能做的乃是证明一种理论较之另一种理论的优越性。在我们关于水沸腾的一系列例子中，我们从未能表明现有理论是真的，但在每一阶段，我们能够表明这种理论较前一种理论更优越。在任何给定的时间里，这是所有科学的一个特征。科学是即定事实的身躯，这种流行观念是完全错误的。在科学中，没有什么东西可以加以永恒地建立，没有什么恒古不变；实际上，科学显然时时刻刻都在变化着，而且这种变化不是通过增长新的确然性进行的。如果我们是有理性的，那么我们总会把我们的决定和期望建立在‘我们的最好知识’基础之上（正如通俗的用语所说的），为了实际目的而暂时假定那种知识的真理性，因为它可能是我们所获得的根据中可靠性最多的一个；但

是我们决不能忘记这样一个事实：它随时都可以被经验表明是错误的，需要我们加以修正。”

“波普尔的‘真理’观很象这种观念：在寻求知识的过程中，我们的兴趣在于越来越接近真理，我们甚至可以知道我们取得了进展，但我们绝不能知道我们是否达到了我们的目标。‘我们不能把科学等同于真理，因为我们认为牛顿和爱因斯坦的理论都属于科学，但是它们不能两者都是真的，也许它们两者正好是错误的。’”

波普尔对传统的西方哲学进行了全面的批判，成功地建立了证伪主义哲学和进化论范式的认识论。但是如同一切科学理论最终都要被证伪一样，波普尔的学说也存在着严重的缺陷。但这种缺陷不是产生于他对西方传统哲学的批判，而是产生于这种批判的不彻底性，他的批判矛头几乎指向整个西方传统哲学，但不可能指向他自己立足的地面——大鱼模型的深层精神结构。

波普尔的证伪主义哲学的重要的缺陷是，他要解决休谟问题的企图并没有成功。证伪主义为科学建立的大陆仍然是不固定的，波普尔仍是一个怀疑论者或不可知论者，在科学不能得到真理这一点上，波普尔与休谟的结论只是五十步与百步之差。休谟认为归纳证明不了真理，而在波普尔的证伪主义体系之中，即使我们手中就掌握着真理，我们也不可能知道这一点。波普尔寄希望于证伪主义能够发现不是真理的理论，把谬误逐步消除就可以使我们逐步接近真理，但是逐步接近真理的目标也是无法实现的。这是因为在实践上我们无法对两个理论判断哪一个更接近真理，甚至无法判断哪一个理论是真理的概率更高；而在理论上，波普尔所建立的进化的、开放的认识论体系认为科学理论并非世界的客观事物的身躯而是人类思维的产物，或用爱因斯坦在信中赞同波普尔的话来说：“理论不会从观察的结果中构造出来，它们只能被发明。”这与波普尔所声明的自己有着形而上学的信仰，例如相信

自然规律的存在之间是明显矛盾的。在一个开放的，充满了发明的认识论体系之中，认识的发散性与逐步接近真理的收敛性是不相容的。如果我们认为真理是客观存在的，是隐藏在变化多端的现象背后的稳定的规律，是一种绝对不变的有限的存在，那么认识的最佳方法应该是一个发现的过程而不是发明的过程。如果我们只能通过发明新的理论来逐步接近真理，那么在这个接近过程之中，新的理论与前一个理论的差异应该越来越小才对，就象数学中的极限概念一样，是一个收敛过程，但是在科学的发展史中并不呈现任何理论上的收敛，从亚里士多德物理学到牛顿力学，从地心说到日心说，从绝对时空观到相对时空观，在理论本体上都发生了翻天覆地的变化。如果我们认为真理是无限的或不断发展的，那么对真理的接近也就成了不可实现的愿望。实际上波普尔理论中的矛盾与休谟问题都起源于同一个错误的假定，即科学和认识的目标在于追求真理。为了化解由于真理的绝对性所引起的矛盾，也有人提出了相对真理或多元真理这样的概念。但是这些剥夺真理的绝对性和唯一性的作法，只能引起真理概念的退化，使真理的概念失去唯一性，永恒性，普适性等优越的性质，变成了与普通的理论并没有什么大不同的东西，失去了真理的魅力。一切企图把真理从绝对化所引起的矛盾中挽救出来的作法，都只会使真理的概念发生退化、模糊化、庸俗化，使真理失去光辉。

波普尔的一个重大贡献在于，他明确指出“真”并不是判断科学理论好坏的唯一标准，作为科学的理论，必须具有可证伪性。一个象大实话或废话的理论最可能是真的，但它不能给我们增加知识，因而不受欢迎。一个科学的理论，应该具有高度的冒险性，从这个理论可以导出很多可检验的推论，这样的理论才称得上是科学的，可证伪的理论。波普尔的学生约翰·沃特金斯（John Watkins）对于科学的目的进行了系统而深刻的论证⁽⁸⁾。沃特金斯首先规定科学的目的应该满足的五个条件：

1. 是一致的，
2. 是可行的，
3. 可指导对竞争理论或假说的选择；
4. 无偏袒，
5. 包含真理观念。

他先提出一个不一定满足上述条件的完全乌托邦式的科学目的；追求(A)确定为真的(B1)终极的(B2)统一的普遍原理，结合关于初始条件的真实描述可以说明或(B3)预见所有的经验现象。这个目的被称为培根——笛卡尔理想。然后逐步修正这个乌托邦式的目的，使其满足上述的五个条件，从而得到沃特金斯认为科学所应追求的最佳目的：

- (A) 可能为真，
- (B1) 越来越深刻，
- (B2) 越来越统一，
- (B3) 越来越有预见力。

在论证最佳目的的过程中，沃特金斯正确地把目的分为两极：由(A)代表的“可靠极”和由三个B所代表的“深度极”，并且指出在两极之间是有冲突的。为了解决这个冲突，“迁就要求(A)的尝试已经致使科学牺牲了一个又一个有价值的特征：说明的深度、普遍性、物理实在论。并且所有这些牺牲都是白白作出的牺牲”，因此必须让(A)作出让步，使可靠极由“确定为真”降低到“可能为真”。

我们已经看到，沃特金斯的关于科学最佳目的的体系是一个追求“真”和“善”的混合体系，四条之中有一条是关于“真”的，三条是关于“善”的，而关于“真”的那一条不断引起理论上的麻烦，被迫不断被消弱，从“确定为真”到“越来越可几”再到“可能为真”。那么我们为什么不再向前迈出关键的一步，干脆取消追求真理的目的，统一到追求“善理”的目的上来呢？沃特金

斯为了保存(A)提出了两点理由,而这两点理由在我看来都是站不住脚的。第一个理由是科学的目标必须满足条件5,即包含真理观点。而对于条件5,沃特金斯只作了这样的论证:“关于第5点,我必须承认,关于这个条件不是完全没有争议的,因为——除非我有误解——这个条件实际上已经至少被当代一个哲学家所摒弃(劳丹,[1977],第24页和第125—126页)。但我认为,如果说真理不是科学目的的一部分,就等于说治愈不是医学目的的一部分或者利润不是商业目的的一部分一样。”在我看来,这个“等于”是不成立的,更好的类比是,如果说真理不是科学目的的一部分,就等于说上帝不是认识的目的的一部分一样。沃特金斯的这个论证除了说明真理的观念是大多数西方人的形而上学的信仰,是上帝在科学中的代替物之外,并不说明别的什么。真理对于科学,就象上帝的第一推动对于牛顿力学一样,只是画蛇添足而已。第二个理由是科学的理论要求验证,而这种要求只能由(A)提出来。为了证明这一点,沃特金斯依靠了对预见力这一概念的歪曲,他认为预见力只是指出预见,而不包括验证预见的后验要求。对预见力的这种看法当然是错误的,如果同意该看法,就意味着我们应该认为一个信口胡柴的算命先生是最有预见力的,从实践的角度来看,按照对“可能为真”的沃特金斯的要求(也是波普尔的证伪主义的要求),科学家应该对自己提出的理论作最大的努力进行以证伪为目标的检验。但在现实之中,大概没有哪个科学家能够忠于证伪主义,热衷于对自己的理论作这种“破坏性试验”,他们只是乐于对自己理论的最精采的预言进行检验。

真理观念的形而上学性是怀疑论的根源。如果我们把追求真理作为科学的目标,把追求真知作为认识的目标,那么必然会陷入大量的绝对性所引起的理论困难之中而不能自拔。只要我们能够象拒绝其他形而上学观念一样地抛弃真理观念,把科学和认识的目标改为追求更好的理论(善理)和知识,那么怀疑论就根本

不会产生，因为只要我们给定一组评价准则（例如沃特金斯的(B1) (B2) (B3)），就可以合于理性地在几个理论中选择最好的一个，而且基于“善理”的体系是一个开放体系，永远不会对将来出现的更好的理论关上大门，只要我们还活着，就足以证明我们已掌握的知识对于维持自身生存来说是够好的。

把科学和认识的目标由“求真”改变到“求善”并不是一件轻而易举的事情，对于习惯于依赖形而上学的信仰的人来说，相对的“善”比起绝对的“真”是一个软弱得多的基础。“真”的绝对性和客观性给人一种非常坚固而可以依靠的印象，同一个理论不可能对某些人来说是真的，对另一些人来说是假的。而“善”的相对性则看起来很不稳固，可以说没有两个人会持有完全相同的评价准则，对于某个人来说的“善”，对于另一个人就可能是“恶”，最终会不会导致科学的理性争论被投票表决所代替？但是这种担心是浮浅的、局部的，也是短视的。从整个科学体系来看，任何一个按照大鱼模型建立起来的理论体系，不管它的绝对的形而上学的基础在一开始看起来多么牢固，都不可能承受理论的无限发展，总有一天是要被压垮的，随着理论的深入发展，绝对性所引起的矛盾和悖论总会暴露出来，导致绝对的形而上学基础的崩溃。科学的发展史，也是一个逐步摈弃形而上学的历史。按照地球模型建立理论体系，在一开始由于理论太小，还象一颗小行星，显得漂乎不定，缺乏引力，建造起来很困难，但随着理论的增长，会逐步稳定下来，而且越来越稳固，引力越来越大。从理论体系的未来来看，地球模型是优越的，但是在理论体系的初期，自立性比较差的时候，形而上学的支持是不可缺少的，因此现代科学只可能产生于一神论的背景，而当科学理论发展到一定程度时，自然会压碎它的形而上学基础，但这时理论体系的自立性已经形成，可以转入地球模型了。所谓“物理学的危机”并不是物理学本身的危机，只是物理学的形而上学基础的危机，两者不应

混为一谈。怀疑论的困难也不是科学本身所遇到的困难，而是关于真理的形而上学观念的困难。这也就是为什么在哲学界出现的形形色色的怀疑论并不会影响科学家研究科学的兴趣。

在现代科学的地球已经建立起来的今天，我们建立一个理论体系时就不再需要强有力的形而上学的基础作为其出发点，只要作一个“锚定”的工作，即把理论起点的基本原理和概念尽可能和现有的科学体系联系起来，用科学的支持代替形而上学信仰的坚定性，这样随着理论的发展和深化，其基础也会越来越牢固，树长得越高，根扎得越深。我们将采用这样的方法建立“善理”的体系，也就是关于评价的基础理论。如果没有这样的理论，“善”将是一种不便于理性思考而只适于投票表决的东西。“善理”是我们在下一节建立非形而上学认识论的基础。

作为善理理论体系（善理论）的起点，我们提出善理的三个基本原则：

1. 存在原则；
2. 经济原则；
3. 预测原则。

这三个原则不象几何公理那样成并列关系，而是有先后顺序的，存在原则是最基本的，从存在原则可以推出经济原则，从经济原则可以推出预测原则，所以相当于公理地位的只是存在原则，另外两个基本原则相当于定理的地位，由于在善理论中具有同等的重要性，把它们列为三大基本原则。下面就分别讨论这三个基本原则。

存在原则：这里所说的存在，是认识论意义上的存在，而不是本体论意义上的存在，因此它既不是唯物主义的客观存在，也不是唯心主义的绝对精神那样的存在，因为这两者都是本体论意义上的存在，而认识论意义上的存在是“祭神如神在”那样的存在。笛卡尔企图用理性的方法去证明本体论意义的存在，从怀疑

一切存在出发，找到唯一的可靠起点是“我思故我在”。但这个论证却不是本体论的而是认识论的，“我在”依存于作为认识主体的“我思”。如果真正追求本体论意义的存在，则只能依靠形而上学的信仰，而最后得到康德的永在彼岸的可望而不可及的“自在之物”。存在原则所说的是，对于某一事物来说，有利于该事物存在的因素对于该事物来说是善的，而不利于该事物存在的因素，对该事物就是恶的。对于习惯于绝对论和本体论的人来说，这种庄子的“此亦一是非，彼亦一是非”的相对主义说明显得缺乏稳定感，但是随着后文中认识论的逐步建立和充实，我们对存在和存在原则的理解会逐步加深，建立起比较完整的存在观。存在的基本条件是稳定，也可以说“存在即稳定”，存在原则也可以称作“稳定原则”。稳定性是在科学领域中被广泛而深入地加以研究的概念，可以说稳定是存在这一哲学概念在科学中的对应概念。建立这种对应关系就是前文所述的“锚定”，是哲学的非形而上学化的一个重要方法。在这里要建立的另一个对应关系是存在与进化论的对应，即达尔文所说的“优胜劣汰，适者生存”的观点，这里的“优”和“适者”是用其生存的可能性来定义的，是一种“存在即善”的观点，是存在原则的一种表达形式。要注意的是认识论意义上的存在是相对的存在，而本体论意义上的存在是绝对的。在本体论意义上，存在与否是一个二值函数，要么存在，要么不存在，没有半存在半不存在的中间状态。而认识论意义上的存在则是相对的，稳定性是相对的概念，适者与非适者也是竞争中的相对表现。这种相对性是善恶所固有的相对性所要求的。站在认识论的立场上来看，本体论意义上的绝对的存在是没有意义的，我们只能讨论相对的存在。例如我们可以研究人和上帝哪一个存在更为稳定。如果我们向全世界的人提问，上帝是否存在和人是否存在，回答上帝存在的人可能不到30%，回答人存在的可达100%，（如果有人回答人不存在，那么等于他宣布自己作为回

答者也不存在，可以不计入统计数字）因此可以说人是比上帝更稳定的存在。这个结论是相对的，如果我们改变调查方法，只去问一个基督教徒，那么会得到另一个结论，上帝在人存在之前就存在，在人类灭绝之后仍然存在，因此在这个人的认识体系之中，上帝是比人更稳定的存在。如果问一个不信任何宗教的人，那么在他的日常生活之中考虑与人有关的概念比考虑与上帝有关的概念频度高得多，那么在这个人的认识体系之中，人是比上帝更稳定的概念。相对性是指对于时间、地点、目的、方法、评价准则等各种条件的依存性，如果去掉所有这些条件，从本体论意义上讨论绝对的存在，那么就只能得出象康德那样的二律背反，用同样的逻辑推理，既可以证明上帝的存在，也可以证明上帝不存在，或者象一般科学界所公认的那样，认为用科学的方法既不能证明上帝的存在，也不能证明上帝不存在。

经济原则：经济原则可以用存在原则加上生物学原理推论出来。可以暂时认为自然的非生物界没有经济原则可言，对于行星的运行来说不需要讨论其经济性。对于生物来说，生存就是作为生物的存在，生物的特征就是要进行新陈代谢，而进行新陈代谢就必然要消耗能量，生物总要通过各种生命活动从外界获得能量或含有能量的物质以克服热力学第二定律的倾向维持低熵状态，由此就产生了经济原则。如果一个生物是“不经济”的，即能量的总支出大于总收入而处于亏损状态，那么生物就不能维持其作为生物的存在。从进化的观点来看，处于相互竞争状态的两种生物，如果一种的经济性能高于另一种，那么即使两种生物开始都不出现赤字，也会由于马尔萨斯原理使经济性能差的生物先出现赤字，或者由于黑字的差别使得经济性能好的生物有更多的能量提供给下一代，从而提高其进化中的适应度，在进化过程中使经济性能差的生物被淘汰掉。虽然经济原则依靠生物的性质而成立，但可以推广到非生物界。首先考虑人工制品的世界，由于人是生

物的一种，遵循经济原则行事，其创造物也就必然体现这一原则。河流本身没有经济可言，但是当它作为人类索取能量的对象而存在时，流量和落差就显示出其经济意义。星球的运动不存在经济问题，但是关于行星运行的理论所产生的计算复杂程度是有经济性的，可以说日心说代替地心说是经济性对神学形而上学性的胜利。站在认识论的立场上，认识主体——作为生物的人的存在是必然的前题。即使想象一个完全由机器人组成的世界，只要机器人能够认识世界，有资格成为认识主体，就必将依靠经济原则，这一点将在后文说明。将经济原则进一步具体化，例如定义出能量的收支比或收支差与时间之比作为经济性的评价参数，就构成了效率原则，效率原则是经济原则的直接产物。经济原则的相对性是不言而喻的，我们难以想象一个绝对最经济的存在。我也没见到过关于经济的本体论意义上的定义，诸如“上帝是经济的”“物质的本质是经济的”之类说法。经济原则可以这样表述：一切生物、作为生物的人、人的认知过程、人利用自然的过程、以及一切具有认知能力的存在都有追求经济性的倾向，经济性是善的重要指标，经济性可以具体化成为效率。

预测原则：预测原则来源于经济原则和人的重要特性，人与动物的一个重要区别是人能够进行有目的的活动，建筑师盖房子与蜜蜂筑巢和蜘蛛结网最大的不同是建筑师在盖房子之前就已经有了图纸，而蜜蜂和蜘蛛没有（你怎么知道蜜蜂没有图纸？回答请参见后文）。预测能力或称预见力是进行有目的的活动的必要条件。预测的失败会导致不经济的结果，甚至危及生存。提高生存概率和增加经济性是人类认知和动物行动模式的根本目标。而对于人类来说，认知与科学活动的更直接的目标是提高预见力。这就是预测原则，是对认知活动和科学理论进行评价的重要原则。当然，人与动物在预测能力方面的区别是相对的，虽然人类表现出比动物高得多的预见力，但这里有一大部分依靠了文化积累而非

生物特性，在生物特性方面人的预见力或提高预见力的认知功能是在漫长的动物进化过程之中逐步积累起来的，因此不能认为动物是完全没有预见力的，人的预见力归根结底来源于动物界。预见力的相对性是显而易见的，既使我们可以象本体论者想象的那样掌握真理，由于对初始状态不可能完全掌握以及误差放大的机制，再加上指数爆炸引起的计算不可能性，都使得绝对完全的预测成为不可能的。

从这三条基本原则的导出过程来看，有一个基本的顺序，从存在原则（稳定原则）到经济原则（效率原则）再到预测原则；其适用范围也逐步缩小，从一切事物到生物，再到人类认知。但是这个顺序只是为了初期建立理论基础的方便而设定的，这三条原则还有相互支持，相互影响的一面。存在原则虽然是对一切事物而言，但是从认识论的角度来看，我们所看到（知道）的一切事物都依存于我们的认知结构，依存于我们的感觉机制和理论框架，特别是依存于我们的分类原则，而这一切认知结构又都是依据经济原则和预测原则以及稳定原则通过生物进化、文化进化和个人认知的过程建立起来的，因此存在原则也必然受到另外两个原则的支持和影响。这里又出现了一个循环论证，但是当我们企图用人的认知能力去认识人的认知能力时，就已经注定了循环的命运，与其拒绝循环而导致理论的形而上学崩溃，或是让循环绕个大圈子再从后门羞答答的放进来而早晚要被人知道，还不如痛痛快快的接受它，堂而皇之的讨论为好。

人类认知的一个方法是建立因果关系，因果关系是在时间先后的事物之间的稳定联系。只有掌握了这种稳定联系，才有可能从前因预料后果，进行有效的预测。因此建立因果关系是人类认知的基本倾向，康德也把因果关系作为认识所必要的先验范畴之一。所谓理论，也就是概念之间的因果关系网。古代的泛灵论世界观，也是追求因果关系的结果，人们看到日月的运行，就想知

道其原因，从人类自身的运动能力出发，推论出日月运行也象人一样由其灵魂引起其运动。这种类比式的推论方式是人类认知的基本方法之一，虽然它不能保证成功，但是可以提供假说。建立因果关系有利于预测，但是即使不明确因果关系，人类对于预测的根本要求依然存在。这就是为什么自古以来形形色色的占卜流行于世的原因，也可以说明具有更强预见力的科学理论为什么代替了其他理论，以及在科学昌明的今日，在科学尚无法给出预测的领域里占卜为什么仍然受到欢迎。

从泛灵论的例子可以看出，并不是所有因果关系都是稳定而有利于预测的，我们有可能建立起貌似有理却缺乏预见力的理论，如果我们认为太阳是由某个灵魂推动的，那么完全可以想象那个灵魂一不高兴太阳在早上就不升起来了。这也就是我们不把因果律作为基本原则的原因。波普尔敏锐而深刻的看到了这一点，提出了以可证伪性作为科学与非科学的划界标准的证伪主义立场。可证伪性与预见力是一致的概念，只有能作出有效预见的理论才是可证伪的，而只有经受住检验的预见才是有效的。对于预见的可检验要求来源于预见力的起因，即有目的活动对预见力的要求。因此对预见力的评价也必然依存于有目的活动的成败的检验。一个理论如果不可检验，则说明它对有目的的活动不发生影响，是没有预见力的。

对预见力的评价可以包括三个方面，即预见的广度、预见的精度及预见的成败。理论的适用泛围越宽，其预见的广度越好，则理论或知识的经济性和稳定性越好。预见的精度及预见的成功则是有目的活动的要求。从证伪主义的角度看，预见的广度相当于理论可以导出更多的可证伪的现象，预见的精度相当于可证伪性，预见的成败相当于理论是否被证伪。预测原则与证伪主义的一致的作用在于排斥两种理论，一种是不可验证的理论，例如“作好事的人死后可以升入天堂”，另一种是永远正确的理论，例如“将

来一定下雨”，即使过了一百年，整个地球一滴雨也没下，也不能说这个预见是失败的，因为将来可以无限延续下去，这种理论是没有内容的。预测原则与证伪主义都重视知识的增殖，但是二者的理论出发点不同。证伪主义的出发点是追求真理，通过知识的增长和证伪过程消去其中非真的部分而逐步接近真理。但是这个出发点会引起矛盾，从追求真理的角度出发，排斥永远正确而没有内容的理论是缺乏充足理由的。所以证伪主义仅仅作为科学与非科学划界的标准，把永远正确的理论逐出科学领域了事。但是如果忠实地执行证伪主义，就难以理解科学领域中保留了很多已被证伪的理论这一事实。例如“某药对某病有效”可以被一个无效病例所证伪。而且证伪主义也不能说明科学为什么比非科学更为合理，似乎喜欢科学仅仅是一种个人的偏爱。预测原则的出发点是追求善理，即追求更好的理论。而好坏的立足点是人类认知的生物学基础。这种观点依存于“人之初，性本善”的深层精神结构，又有其进化生物学的理由，即经过几十亿年进化淘汰而产生的人类，其生物本能所规定的价值观和评价体系应该可以说是久经考验的，即历经反复验证成功的。人的价值观有来源于生物进化的，也有产生于文化进化的，而生物性价值观比起文化性价值观要稳定得多，从时间轴上看，生物进化比文化进化缓慢得多，从空间轴上看，全世界的人都属于同一生物种而分别属于不同的文化集团，因此不论从时间还是空间来看，生物价值观都比文化价值观具有更高的稳定性和良好的一致性，作为评价理论体系的基础更为可靠。从追求善理出发，预测原则等三大原则并不在科学与非科学之间划界，而是可以分析科学领域中通行的价值观比其他价值观更符合人类认知的生物基本要求，从而说明在科学与其母体一神论宗教发生冲突时为什么科学能够取胜，也可以说明为什么现代科学可以轻而易举地跨越文化壁垒在全世界畅行无阻。当西方文化大举进入中国的时候，现代科学广为人们所接受，

而基督教却在“子不语”面前碰壁。从善理的角度看，科学是人类认知发展的高级阶段，科学哲学与认识论应该是一致的，科学的理论与其他理论都是人类认知的产物和工具，区别仅仅是从以生物学为基础的三大原则来看，科学理论优于其他理论而已，由于优劣的相对性，科学与非科学之间不应有绝对的划界。

站在善理的立场上建立理论实际上是要求我们建立关于价值观的理论体系。为建立这种善理的理论体系，我们可以采用科学领域中最通用的欧几里得范式，从一些最简单的公理出发，推出其后的众多定理，从而构成一个开放体系。善理的三大原则就类似于几何公理处于理论体系的起始位置。有区别的是，这三大原则不是形而上学的起点，而是可以用科学的方法加以论证的，或者说是“锚定”于科学地球之上的。欧氏几何中定理的证明，依靠的是演绎方法，即逻辑上的一致性的保证。价值观体系中价值的论证，依赖于价值的一致性。其基本思路是，对于某种价值观，我们可以论证它与基本价值观或已经论证为合理的价值观之间的关系，如果是一致的，或者说提高其价值有利于提高基本价值或其他合理价值，那么这个被论证的价值观就是合理的。如果能够建立某种价值与其他已知合理价值之间的评价函数关系，那么这种一致性的证明就比较严密，当这种函数关系并不明确时，只要我们能够找到其一致或相反的增长方向，论证也可以进行。演绎逻辑依靠二值代数，逻辑函数的取值只有真和伪，0和1，而评价函数往往是连续函数，这是与演绎逻辑不同之处。与欧氏几何相类似的是，当我们论证某种新的（尚未纳入体系的）价值观时，往往也需要新的定义和一些技巧，例如如果不引进三角形和圆的定义，就不可能有关于三角形和圆的定理，几何证明需要技巧是众所周知的。

我们在前面建立善理的三大原则时就使用了这种手法，先建立存在原则，然后把领域局限于生物界，定义出对于生物来说的

存在就是生存，再根据生物的新陈代谢和能量守恒导出收支关系的经济原则；进而把领域局限于人类，找出人类有目的活动的特征，根据预测的成败对经济性的影响，论证预测原则与经济原则在评价上的一致性；即预测能力有利于人作为生物的经济性，而经济性有利于生物的存在。

用这种方法可以分析和论证在科学领域中一些价值观的合理性。当然不是说非科学领域的价值观例如宗教价值观都是不合理的，但其中有一些是不合理的，有一些则分析起来比较复杂，要到本书后面再作分析。而科学领域中的价值观与三大原则符合得较好，因此在这里先只讨论科学领域的价值观。

前面已经分析过为什么波普尔的逐步接近真理和沃特金斯的“可能为真”不是科学所应追求的目标，“可能为真”的概念应该用预测的成败来取代。下面要分析为什么深刻的、统一的、有预见力的理论是科学所应该追求的目标，即合理的价值观。沃特金斯所采取的论证方法是先提出一个可能想象的最高目标，即所谓的培根——笛卡尔理想：要求科学对所有已知现象提供一个统一的和终极的说明，这种说明是确定地真的，使所有未来现象都完全可预见。然后按可行性的要求把这个乌托邦式的理想降为现实的目标，即把追求统一、终极、确定为真、完全可预见的理论降低为追求更统一、更深刻、可能为真、更有预见力的理论。在背后支撑着培根——笛卡尔理想的，实际上是一神论的深层精神结构。如果我们相信是唯一的上帝创造了世界，那么上帝创世时所依据的唯一的理论就是我们所追求的统一的、终极的、确定为真的具有完全预见性的理论，这也就是隐藏在纷纭变幻的现象背后的最深层的永恒不变的规律，是自然界的身躯。但是对于不具备一神论深层精神结构的人来说，沃特金斯的论证就失去了根基，我们需要从另外的出发点即善理的三大原则来讨论统一、深刻、有预见力的理论为什么是值得追求的，为什么是我们有理由接受的

价值观。为此我们先用图 2.3 表表现在认知过程中理论或概念与现象之间的关系模型。图中的方块代表现象，圆圈代表概念和理论，连线代表解释和支持的关系。从这个图中我们可以看出通常所说的理论的“深度”的含义。越是深刻的理论，它所能解释和预见的现象也越多，而且相对于浅层理论来说其理论的数量较少，这既增加预测现象的广度，也符合经济原则。图中的虚线代表未知的理论与现象，理论越是深刻，其增殖出新的表层理论和预测未知现象的能力也就越大，这也是与预测原则和经济原则一致的。另一方面，理论越是深刻，离现象也越远，由于受部分现象变化而导致理论失效的可能性就越小，那么理论就越是稳定，这也符合稳定原则。

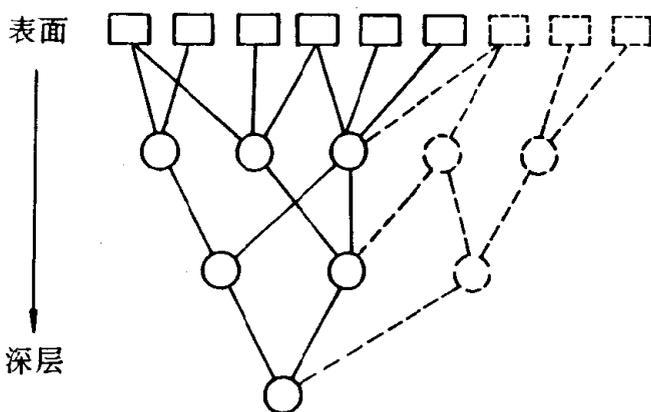


图 2.3

关于理论的统一性要求，可以分为图 2.4 中的两种情况来讨论，*a* 表示两种不同的理论分管不同现象且现象之间互不相干的情况，就如开普勒的物理学管天上的行星运动而伽利略的物理学管地上的落体运动那样。但是这样的理论上的分裂局面总是给人一种深度不足的感觉，我们会期待图 *a* 中虚线所表示的更深刻的理论把两种理论统一起来，因此在牛顿力学出现时受到广泛的拥护。分裂的理论总会在领域的交接处出现困难，例如我们可以想

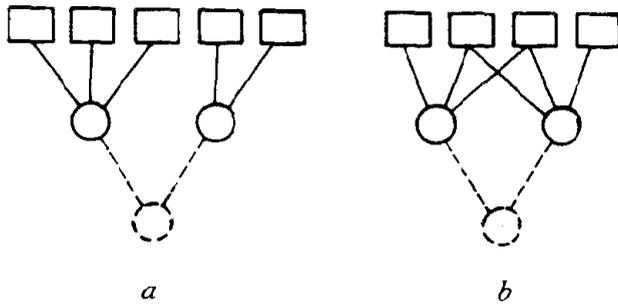


图 2.4

象这样一个问题，如果一个物体由地升空，逐渐远离地球，那么是否存在一个临界点，在这之前物体遵循伽利略定律，而在其后按开普勒的理论行动，开普勒和伽利略对此都不能回答。这个处于理论边缘的物体把图 2.2 中 *a* 的情况变成了 *b* 的情况，造成更大的理论困难。*b* 的情况是指两种不同的理论都可以解释一些共同现象的情况，典型的例子是中医和西医的理论完全属于不同体系却面对相同的患者。两种理论优劣互见，让人无所适从，如果有一种更深刻的理论把两者统一起来，从症状可以判断该看中医还是西医，显然可以避免治疗过程中不经济的做法。但是现在还没有这样一种理论，我们只好靠经验先选定一种，不灵了再试另一种，造成很大浪费。从理论上讲，分裂的理论面对共同现象时应该可以通过检验来排斥其中的一个，但是现在任何的医学理论都还没有达到足够的硬度，即预测的精度和成功率都不象物理学理论那样高，所以不经济的作法还难以克服。目前关于中西医结合的理论研究主要倾向于用西医理论去解释中医理论，也就是以西医理论来统一中医理论。这是由于西医理论有一个明显的优点，就是与现代科学的其他领域，例如物理学、化学、生物学之间的一致性较好，更为统一，而中医理论对于象 CT、B 超这种现代检查方法则难以产生关联。因此西医中药这种方法成为中西医结合的主流。不管在何种情况下，理论的分裂局面总是暴露出理论的

肤浅和不完善，是令人不能满意的情况，是违背三大原则的，因此统一性对于科学理论体系来说是很重要的。从科学理论的“地球模型”的角度看，理论的统一性是组成大理论体系的前题，其价值也是显而易见的。

与理论的统一性密切相关的是理论的内部一致性，即为什么一个自相矛盾的理论不能被人们接受。当我们用理论来预测现象时，一个自相矛盾的理论会导致在不同的推论路径下，从相同的理论和前题推出相反的预见，使我们无所适从，这样的理论显然是没有预见力的，不符合预测原则。演绎逻辑是保证理论内部一致性的重要工具，因此在善理体系之中对演绎逻辑给以充分的评价。但是站在善理的立场上看，演绎逻辑既不保证理论的真理性，也不是思维的深层机制（象人工智能的逻辑主义者所想象的那样），更不是构成宇宙的框架，它仅仅是在文化进化中被创造出来的工具而已。当然，演绎逻辑作为理论工具的重要性是不言而喻的。

科学是追求客观性的。按照通常的方法我们可以对此给出一个形而上学的解释，即这是由于科学家无论他们的信仰如何，在从事科学事业时他们必须作为一个唯物主义者来活动，即相信客观世界是脱离于自身而独立存在的，相信客观世界是有规律的，如果没有这种最起码的形而上学的信仰，科学就无从开始，虽然我们无论怎样努力，都难以对这一形而上学的信仰给出合乎理性的证明，培根的归纳法的合理性是不能用归纳法证明的。从善理的角度出发，我们并不需要对客观性的价值作形而上学的解释。客观性意味着一种稳定性，对客观性的追求与对理论或现象的稳定性的追求是一致的。通常我们认为客观的东西，都是具有某种稳定性的东西，如同我们眼前的桌子，不管用视觉还是触觉，都能感到其存在形式，而且这些感觉之间还存在着稳定的相关关系，不仅我们自己知道如果没人把它搬走，明天桌子也还在这个地方。而且，如果问一问面对这张桌子的其他人，也会得到相同的结论。这

种稳定的可预测性就是客观性的认识根源。因此对于那些具有相似的稳定性的坚实的理论，我们容易把它们想象成客观世界背后的身驱。破坏一个桌子很容易，破坏牛顿力学要困难得多。有不少数学家认为数学定理是客观存在的，虽然在普通人看来数学法则的客观性远不如桌子那么直观，这是由于数学家们的实践使他们能感到破坏一个数学定理比破坏一个桌子难得多。对于那些稳定性较差的东西，例如难以再现的梦境和想象，难以获得广泛支持的个人看法，与其他感觉和知识框架发生冲突的错觉，都会被我们看成是主观的东西。因此科学对客观性的追求，实际上是对稳定性的追求，是符合存在原则即稳定原则的。

在科学中关于实验有一个与客观性密切相关的价值观，就是对科学实验的可重复性要求。可重复性和不变性一样是稳定性的一种形式。如果一个已经发表的科学实验结果不能被他人成功地重复，一种可能性是这个实验本身面对着不稳定的现象，或者说实验设计不良而没能抓住稳定因素和有效地排斥不稳定因素，那么这些现象即实验结果对于建立稳定的有预见力的理论是没有多大价值的；另一种可能性是实验中出现差错或原实验者蓄意伪造实验数据。科学家的基本行为准则是“怎么干都行”，但是唯一被禁止的事就是伪造实验数据，这是科学界最重视的职业道德。一个科学家是个酒鬼或色狼并不影响他在科学界的声誉，大众传媒对于政治家或明星的私生活津津乐道，对科学家的隐私则不感兴趣。但是一个科学家如果伪造实验数据则会在科学界身败名裂。按照形而上学的解释，科学的目标是追求真理，因些不允许有伪造的东西混入。但是按照善理的价值观分析，一个科学家如果提出荒唐的理论、错误的证明，或是对实验结果作牵强附会或夸大的解释，都不违反职业道德。因为这些东西都是可以一眼看穿的，不会造成什么实际的危害。但是科学实验往往要花费大量的时间、精力和金钱，而重复他人的实验又是个不令人感兴趣的工作，所以

通常对于实验结果都先假定它是真实的。在这种情况下，伪造的实验结果往往不容易很快被发现，而把其后的研究者引入歧途，虽然最终是会真相大白，但已经造成研究过程中极大的浪费，严重违反了经济原则。因此不伪造实验数据这一职业道德是有明显经济价值的。当然，科学成果的可靠性并不依赖于科学家的道德良心，而在于它不断地在使用中被检验。因此假如突然冒出一个历史学家，宣称经过他严密考证的结果，牛顿是一个诈骗惯犯，我们对于牛顿物理学的看法也不会有丝毫变化。这是因为科学是根据其可检验性维持其可靠性的，无需依据任何个人权威和偶像，任何偶像的破坏都不会对科学造成影响。相反，在非科学的领域，由于可检验性较差，不得不在不同程度上依赖偶像的卡利斯马性来支持其理论。因此，在科学论证中，以圣人、名人、伟人 or 权威的话作为论据的论证是不能成立的。反过来说，凡是含有“由于是某某人说的，所以是对的”这种逻辑的论证，都不能算是科学的论证。

简单性也是科学理论所追求的价值之一。如果我们说“自然是简单的”或“真理是简单的”，那么这是一种形而上学的解释。简单的理论易于理解、学习和传播，也便于使用，因此很明显地符合经济原则。简单的理论由于涉及的假定或参量较少，由于某些假定不被满足或参量的变化而引起预测失败的可能性也少，因而更加稳定。沃特金斯正确地论证了简单性暗含于深刻、统一和预见力的要求之中，这也就是说简单性与这些要求具有一致的价值取向，因而也是符合三大原则的。

从善理的角度看，形而上学可以看成是将一些价值观推向极端的绝对化产物。而这种绝对化则会造成理论的困境。因此在一个以这种片面绝对化的形而上学体系作为出发点的认识论者看来，科学家“必定象一个肆无忌惮的机会主义者”那样行事。我们在这里不需要为机会主义翻案正名，机会主义是明显违反稳定

原则的。如果从善理的三大原则出发，我们可以看出科学家既不是机会主义者，也不是在两个形而上学的极端之间走钢丝。而是在一个具有高度内部一致性的价值观体系的指导之下行动。这种价值观体系立足于人类共同的生物学性质，因此是高度稳定而不受文化背景影响的。

通过这一节我们为价值观这种颇具主观色彩的概念建立起比较稳定因而更具客观性的理论基础和理论框架，从而使我们看到在崇尚客观性的科学领域中研究主观世界的可能性。这个善理理论体系的基础就是三大原则，其框架就是价值的一致性论证。为了使这个理论体系更加充实和严密，还需要进一步研究很多更具体的评价函数，但是这已超出本书作为非专业人员阅读的范围，而且建立评价函数的工作在科学领域中是司空见惯的事，对于有关的专业人员来说也不需要再多说什么。

在这一节中我们对于善理的价值理论体系主要还是按欧几里德范式建立的，但是理论的进一步展开则需要更为动态的进化范式（或叫作达尔文范式），例如研究评价准则怎样在进化或学习的过程中被创造、被改变和逐步形成系统，这只能在后几章中讨论。但是建立与进化有关的理论又必需以评价为基础，因为进化的机制是变异与选择，而选择是必须以评价作为根据的。这里又出现了自相缠绕的现象，自相缠绕是本书的一个特征，希望读者逐步习惯起来。

维纳的《控制论》是一部划时代的著作。四十多年人工智能的发展都起源于“人是机器”这样一个基本信念，而《控制论》正是为此奠定了坚实的理论基础。《控制论》的最大贡献在于生物与非生物之间的共性，即它们都遵循存在即稳定的原则，从而为我们用计算机来研究思维和认知找到了充足的理由。更重要的是，稳定性通过负反馈的概念得到了严密的科学性，成为可以进行精密计算的对象。二十年前“反馈”这个词还是一个非常专业性的词汇，

现在却频频出现于大众传播媒介之中。在《控制论》中维纳还提到了“情调调节器”和“预测”的概念，从我本人来说对于评价和预测的兴趣也起源于六十年代阅读《控制论》的时候。但是由于这些概念缺乏象稳定性那样的精密科学的支持，因此在人工智能的研究中未能得到应有的重视，也没有被提到哲学高度来考虑，因而阻碍了人工智能的发展。稳定原则为我们打破了生物与非生物的界限，清除了改头换面的各种“活力论”，是思想界的一个重大突破。这一思想已相当深入人心，当然也并不是被所有人接受的。我也看到过某些学者发表这样的言论：“用蛋白质或神经元为元件可以造出能思维的机器，但是用硅为原料则不可能。”持这种看法的人仍停留在《控制论》之前的年代。

维纳为我们打破了生物与非生物的界限，也产生了一点副作用，那就是使我们容易忽视生物与非生物的区别。为了继续前进，我们不得不重视生物所具有的超越非生物的原则，以及人所具有的超越其他生物的原则，这是我们提出三大原则的更为直接的原因。但是这三大原则并不只适用于认知科学，因此在导出三大原则时我们采用了更有普遍性的论证方法。对于认知科学来说相当重要的一点是三大原则使我们得以跳出维纳的如来佛的手心。

建立关于价值观的理论体系，使我们为下一节中建立的认识论模型提供了必要的基础。

2.4 认识论的逆镜模型

为了理解认识论的逆镜模型，先要说明一下认识论的正镜模型，即通常所说的反映论的模型。这个模型示于图 2.5。图中的左半部分相当于图 2.3 旋转了 90 度。方块表示我们能够直接观察到的客观现象，圆圈则表示这些现象背后隐藏着的客观规律，或称为自然的身躯。连线代表这些自然规律与现象的关系，即现象总是

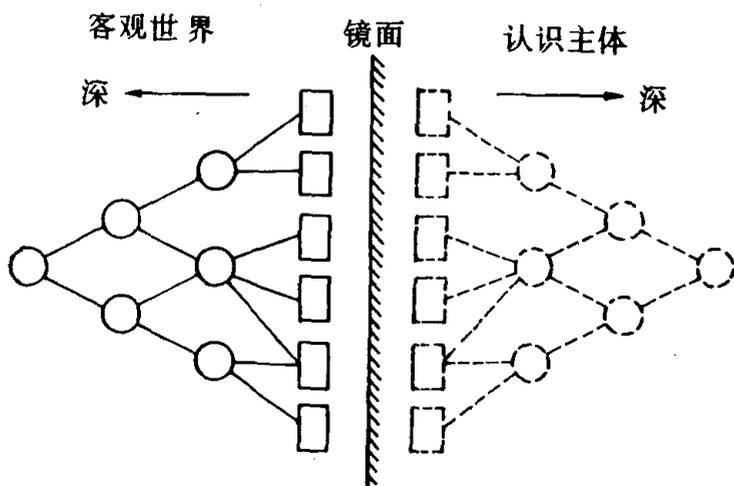


图 2.5

受客观规律支配的。这些自然规律又有深层与浅层之分，浅层的规律受深层的规律支配。为了表示主观认识与客观实在之间的“反映”关系，在主观世界与客观世界之间放了一面镜子。按照一般几何光学的习惯，画斜线的一方是镜子里面，这个镜子面对客观世界。在这个模型中，客观世界是实在的，在图 2.5 中用实线表示，而主观的认识则是客观实在的“映象”，相当于光学中平面镜里的“虚象”，在图中用虚线来表示。主观部分中虚线的方块表示从现象中直接得到的感觉，而虚线的圆圈则代表头脑中的概念和理论。连接它们的虚线则表示理论对现象的解释关系，以及深层理论对浅层理论的解释关系。如果我们的认识之镜是个理想的平面镜，那么镜子里面的虚象将和镜子面对的现实是完全一致的，这相当于理想的认识状态，即我们头脑中的理论与支持客观世界的规律是完全一致的，也就是说我们获得了真理。这就是我们通常所说的我们的思想符合客观规律的含义。

但是，这个被一般常识所接受的认识论模型如果认真地加以

分析和研究,就会发现其中包含着很多引起理论困难的假定前题。

首先,在客观世界中是否象我们想象的那样存在着那种支配着现象的客观规律。如果我们相信上帝按照他的合乎理性的方案创造了整个宇宙的话,那么客观世界中存在着规律这一假定就是合理的,否则这个假定就仅仅是一种不可证明的形而上学的信仰,是缺乏科学根据的。根据“丑小鸭定理”,在完全脱离主观评价的纯客观世界中,不存在任何分类所依据的准则。这就是说,纯客观的世界是一个混沌的世界,是无法进行分类的,而分类又是一切规律所必须依据的前题。因此在一个脱离认识主体的主观评价的无分类混沌状态的纯客观世界中存在着规律或某种客观成立的理论是不可想象的。这就象“在人类诞生之前是否有动物存在?”这样的问题只有在人类诞生之后才有意义一样,如果真的在人类诞生之前又有谁能提出这个问题呢?正如我们拿到一枚古钱,上面印着 256. B. C. (公元前 256 年)我们立刻可以断定这是个伪币。

反映论模型所依据的第二个假定是我们的感觉不会欺骗我们。在理想的平面镜中,客观的现象如实地反映到了主观的感觉之中,虚线的方块与实线的方块是完全一样的。关于感觉的真实性假定显然站不住脚,心理学的实验已经向我们揭示了大量的错觉实例。脑神经科学则告诉我们,各种感觉得到的神经脉冲信息在进入大脑之前和之后都受到了肆无忌惮的加工(或者叫作歪曲),这些信息被变换、被分割、被取舍、被重组,这些处理不管是按什么原则进行,至少有一点可以肯定,即这些处理不是按照“高保真”的原则进行的。彩色电视机能够设计成功,正是巧妙地利用了人的视觉“缺陷”,电视里的黄色可能与生活中的黄色具有完全不同的光谱成分,但我们甘心受骗而不会感到不快。

反映论模型隐含的第三个假定是我们的认识之镜具有光学镜子所不具备的穿透力。我们不仅可以正确地感觉到现象,而且可以透过现象看到现象背后的本质,这样才有可能在认识主体中

“反映”客观规律，才有可能发现真理。也就是说，图 2.5 中虚线的部分才有可能与实线的部分保持一致。使得这个假说成立，是各种认识论都曾面临的最困难的理论问题。最简单的方法，是干脆假定人的认识之镜确实具有这种穿透力，可以直接看透现象背后的规律，感性认识可以自动地产生“飞跃”，变成理性认识。但是这种假定是令人难以接受的。如果我们假定所有人的认识之镜都有这种穿透力和飞跃能力，那么在人类诞生之初就应该已经掌握了真理，这显然不符合事实；如果我们假定只有极少数的圣人、贤人才具有这种能力，那么我们就把人类认识世界的能力建立在极其偶然的突然变异的基础之上，而且难以解释为什么圣人的看法因人而异。这种假定就象活力论的“生命力”假定一样，几乎不能引出任何新的理论内容。脑神经科学、计算机科学、逻辑学等自然科学也都不能对这种穿透力提供支持。因此，大多数的认识论理论都不假定这种直接的穿透，而企图用从直接的感觉之中抽取背后的规律或本质。围绕着间接得到真理的方法，从古至今存在着经验论和先验论之争，但两种理论都存在着难以克服的缺陷。经验论主张通过感觉的重复和联合，可以建立起正确的认识。人工神经网络的研究正是立足于这样的哲学认识论基础。但是人工神经网络模型的研究也揭示了经验论的缺陷。第一个问题是，人工神经网络对付简单问题是可行的，但是随着问题复杂程度的增加，解题的运算呈指数爆炸的增加，单靠经验就不可能得到复杂和高度的认识。这是个工程可行性的困难。第二个困难更有哲学意味，也就是即使我们可以得到一种认识，也不能自动地保证这种认识与客观规律是一致的。也就是说不能保证图 2.3 中的虚线与实线是恰成镜象关系。这是因为从原理上讲对于任何有穷个的现象或感觉都会有无穷多的理论或解释与之符合。这是莱布尼兹早已从逻辑上加以证明了的。我们可以用一个简单的例子来说明这一点。我们通过观察得到的数据在纸上画出有限个点，然后我

们用一条曲线或函数来表达所要得到的规律。很显然通过这些点的曲线有无穷多条。这种局面相当于数学中的反问题。在这种情况下，我们通常是人为的设定一个评价准则，从中选出最佳的曲线，例如最光滑的或符合某个先验理论的曲线。这是工程上常见的作法，但在哲学上引起困难，因为无法解决我们这种主观选择能否保证与客观规律相一致的问题。主客观一致的问题对于先验论来说也同样是困难的。先验论假定我们的头脑中事先存在着某种先验的知识或认知结构。这一点确实是难以否认的。问题在于这个先验的认知结构是什么，以及它是怎样保证主客观一致的。黑格尔的想法是我们的大脑与宇宙的运行包含着相同的逻辑结构。因此可以保证主客观的一致。这个想法隐含着一神论的形而上学支持，需要假定上帝按照这个逻辑创造了世界，而且在创造人的时候把这个逻辑放入了人的头脑之中。也就是假定我们的头脑中持有上帝的一部分蓝图。这个假定对于不具备一神论深层精神结构的人来说是难以接受的。康德对先验论和经验论进行了综合，并且提出了时间、空间、因果性等一些先验范畴作为认知的根据。但是康德自己也并不认为这些先验范畴可以保证主客观的一致性，因而不得不制造存在于认识之彼岸的“自在之物”的概念。

正镜模型是一个基于本体论的包含真理概念的模型，不论这个本体论是唯心主义的还是唯物主义的，真理的概念都会带来类似的理论困难，即使是波普尔或皮亚杰这种进化形式的认识论，只要不放弃真理的概念，都无法逃避主客观一致的义务。休谟问题就是这种理论困境的集中表现。我们通常认为，预测和对预测的检验是有助于检验真理性的。例如通过已知的有限个点我们可以任意画出无限条曲线，可是要让画出的曲线使下一个实验数据点也正好落在这条曲线上则要困难得多。但是从理论上讲，使得以后的有限个点正好符合的曲线仍有无限多条，也就是说预测检验也不能作为唯一性的保证，因此不能作为真理的合乎逻辑的证明。

正镜模型所带来的一连串的理论困难，迫使我们从根本上重建认识论的理论体系，即建立认识论的逆镜模型。为了更好地理解逆镜模型，我们先介绍一个笛卡尔妖魔的概念。有些读者可能对谈科学而使用妖魔的作法感到奇怪，实际上这是在科学中进行思考试验的常套手段。例如在物理学中就有过两个著名的妖魔——麦克斯韦妖魔和拉普拉斯妖魔。麦克斯韦妖魔是这样一个思考试验：一个密封的装有气体的箱子，中间用隔板将箱子隔成两部分，隔板上开一个可以开关的小窗，由一个极小的小妖即麦克斯韦妖魔管理这个小窗，小妖随时观察着向小窗飞过来的气体分子，如果隔板右方的分子高速飞向小窗，小妖就打开小窗把这个高速分子放到左边去，如果右边飞来的是低速分子，就关上小窗把它挡回去，对于左方飞向小窗的分子则采取相反的政策，只让低速分子通过小窗到右边去，而把高速的留在左方。如果有了麦克斯韦妖魔，箱子将自动地左边变热而右边变冷，这意味着违反热力学第二定律，制造出第二类永动机。麦克斯韦妖魔是不可能实现的，这是可以从量子力学得到的结论。但是麦克斯韦妖魔体现了人类合乎经济原则的热烈追求。

拉普拉斯妖魔是物理决定论的一个理想化表现。在牛顿力学中，一个力学系统在某一时刻的各质点的位置和动量将完全决定其后的全部过程。拉普拉斯妖魔是假设有一个神通广大的妖魔，他可以掌握整个宇宙所有质点现在的位置和动量，那么这个妖魔就可以准确地预测整个宇宙未来的全部事件。拉普拉斯把物理决定论提高到哲学的高度，在拉普拉斯决定论的世界中，主观能动性或自由意志之类的概念是没有容身之地的。量子力学的测不准原理和混沌理论的误差放大过程对于物理决定论来说是深刻的问题，但在哲学决定论的世界中引进随机因素也不能拯救自由意志。这个问题对于认知科学来说相当重要，后面还要继续探讨。拉普拉斯妖魔体现了人类对于预见力的强烈愿望。

举这两个妖魔的例子是为了说明在科学领域引进妖魔的合法性，下面我们就来说明笛卡尔妖魔的概念。在道格拉斯·R·霍夫施塔特和丹尼尔·C·丹尼特合著的《心我论》中文版（上海译文出版社）的《译者的话》中有这样一段：“法国哲学家笛卡尔在1641年曾设想过一个十分著名的思维实验。他这样问自己：‘我怎么知道自己没有在受一个魔鬼的欺骗呢？这个魔术通天的魔鬼想引诱我去相信存在着外部世界（和我自己的身体）。’也许这位说过‘我思故我在’这句名言的笛卡尔认为除了魔鬼之外，唯一存在的东西就是他那不朽的灵魂了。这种怀疑主义的论点至今还在流行。随着当代医学科学的进步，这个古老的思维实验又有了现代的说法：我怎么知道某些心狠手辣的医生没有趁我熟睡之际切除了我的大脑？这些居心叵测的医生把我可怜的大脑贮存在一个生命维持系统内，用各种各样假的刺激来戏弄、欺骗我的大脑。”笛卡尔的这个想法写在他的《形而上学的沉思》里。我们称笛卡尔所设想的这个魔鬼为“笛卡尔妖魔”。我们可以想象出一大堆技术上的理由来论证“笛卡尔妖魔”在物理上的不可能性，但是正如物理的不可能并不影响拉普拉斯决定论对自由意志的否定一样，技术上的困难并不损害笛卡尔妖魔的哲学意义。即使从技术可行性的角度看，笛卡尔妖魔所遇到的困难也比拉普拉斯妖魔少得多。例如对于拉普拉斯妖魔要求了解宇宙中所有质点的状态，那么就会出现一个问题，拉普拉斯妖魔自身是否由宇宙中的质点构成呢？如果回答“是”，那么拉普拉斯妖魔就面临着了解自身所有质点状态的问题，这样就会引起无限退行。如果回答“不是”，那么或者拉普拉斯妖魔活动于宇宙之外，这与宇宙的无所不包的定义相矛盾；或者拉普拉斯妖魔只能是虚幻的存在，这也就否定了它的现实可行性。这个问题对于笛卡尔妖魔并不存在，笛卡尔妖魔不需要知道宇宙的一切，甚至可以一无所知。量子力学的困难和混沌现象对于笛卡尔妖魔来说也不是问题。一个有权一手遮天的骗子并不

需要太多的智慧，何况这个天并不大，只包括一个认识主体的所有输出输入信息。因此比起拉普拉斯妖魔，笛卡尔妖魔的神通可以小得多。从我们可以精确地预测日蚀和月蚀可以看出，我们自己在某些领域可以算得上是个准拉普拉斯妖魔。计算机科学中近几年发展起来的“人工现实感”（或称“虚拟现实感”）是我们已经实现的准笛卡尔妖魔。水平更低的准笛卡尔妖魔可以举电子游戏机为例，每当我看到聚精会神地玩电子游戏机的人，我都不免产生一个疑问：到底是他在玩电子游戏，还是编游戏程序的人通过自己造的笛卡尔妖魔在玩他呢？推而广之，我们非常投入地看电影、看电视剧、看小说，不也是暂时地沉浸在笛卡尔妖魔的怀抱里吗？如果在看小说时我们随时提醒自己：“这只不过是印在纸上的铅字”，“这些故事都是作家臭编的”，那还不如不看了。再往大里看，文化环境本身就是一个巨大的笛卡尔妖魔。电影总要散场，小说总要读完，我们总能回到现实之中，如梦方醒，想起自己该吃点夜宵了，一切故事都不是现实。但文化之梦伴随我们从生到死，醒的机会不多，所以这种梦不易被察觉，它的真实性很少被怀疑。只有当你进入另一个文化圈时，才能发现自己的文化。例如你到了外国，语言不通，生活习惯格格不入，你认为理所当然的事，在这里是稀奇古怪，你认为稀奇古怪的事，在这里理所当然。半年或一年之后，你开始用外语说梦话，几年之后，你的言谈举止和当地人没有区别，你用外语和外国人的逻辑思考问题，你已经被另一种文化浸透了。可是当你再回中国，飞机一着陆，周围的景物和空气的气味会唤醒你的一切原有的文化适应，你如梦方醒，完全恢复了一个中国人的所有反应，再回想在国外的生活则恍若隔世。如果其后你也频繁往来于两国之间，头脑中的文化开关可以迅速切换，在外国时觉得中国在梦中，在中国时则外国如梦，无法回答哪个是真正的现实，就象庄子一样，弄不清自己到底是蝴蝶还是庄子，你也弄不清自己到底是中国人还是外国人。

这时你也就体会到了有两个不同的笛卡尔妖魔时的感觉。

目前人工现实感的研究发展很快，也有人在开发使用人工现实感的电子游戏，或将人工现实感用于商品宣传以及更有身临其境感觉的模拟训练装置。但是本书所感兴趣的不是人工现实感能在多大程度上接近真实的现实感这种技术问题，而是在人的认知过程中现实感从何而来，即为什么我们会感到客观世界是实实在在的。人工现实感或笛卡尔妖魔的哲学意义在于，我们感觉到的实在并不证明本体论意义上的实在。人工现实感是否完善或笛卡尔妖魔是否足够神通广大并不重要，关键在于它能否长期稳定，一手遮天。“代沟”现象可以看成为分布于时间轴上的两个不同的文化笛卡尔妖魔在作怪。我们可以进行这样一个思考试验：一个婴儿从一出生起就被放在一个人工现实感的装置中养大，即使这个装置不太完善，他也不会发现他不是生存于真实的世界中，只要我们能够保证他的全部感觉来自于这个装置。

只要我们放弃本体论的存在观而接受认识论的存在观，那么笛卡尔妖魔对我们不会造成任何威胁。认识论的存在观认为，存在是认识主体对于其输入及输出信息进行加工而建立的稳定、经济、可预测的关系。这个说法过于精炼了，还需要加一些解释才易于理解。图 2.6 是一个示意图，说明认识主体与客体的关系只依赖于感觉输入与动作输出。我们不需要假定在认识客体中事先存在的本体论意义上的客观真理或客观规律，也不需要假定认识主体与客体之间的先定和谐。对于认识客体我们只要求它提供动作与感觉之间以及时间先后的感觉之间的相对稳定的关系。至于这个关系是由本体论意义上的客观实在引起的，还是由笛卡尔妖魔制造的，对于认识主体来说，既不可能知道，也没必要知道。认识主体所需要做的和可能做的，只是对输出输入信息按照三大原则进行加工，制造出实在的感觉和深刻的理论。我们所看到的客观实在，只能是我们头脑中由信息加工而产生的世界象，对这个

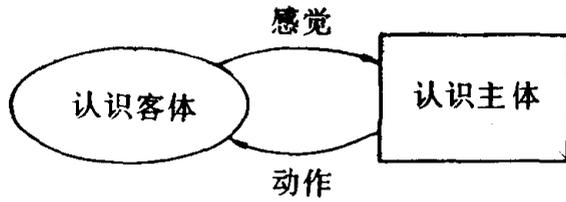


图 2.6

世界象的要求不是“真实”，而是有利于认识主体的存在。生物学的研究表明，青蛙眼里的世界与我们眼里的世界是不同的，争论二者哪个更“真”没有多大意义，重要的是我们的世界象有利于我们生存，而青蛙的世界象有利于青蛙生存。

按照非本体论的认识论观点所建立的认识论模型是一个逆镜型，如图 2.7 所示，它相当于把图 2.5 的镜面翻了一个面，即镜面是朝向认识主体的。称之为逆镜模型是与图 2.5 的本体论的认识论正镜模型相对而言的。图 2.7 中实线的方块代表由认识主体的输入输出按三大原则建立起来的具有现实感的现象模型，圆圈

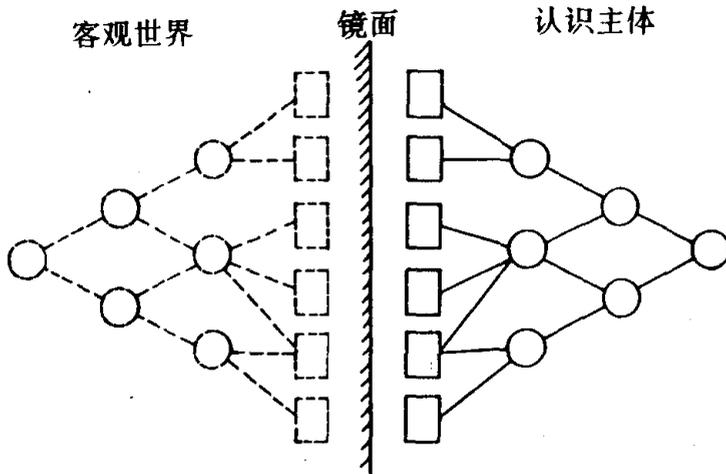


图 2.7

则代表关于客观的深层规律的理论，连线表示现象与理论以及理论与理论之间的关系。认识主体之中的这些认识结果的总和构成了我们的世界象。我们所看到的“客观世界”是这个世界象向认识客体的方向的投影，即图中由虚线所构成的部分。也就是说，认识主体中的世界象是实在的，而所谓的“客观世界”是这个世界象映在镜面背后的虚象。

认识论的逆镜模型完全避免了主客观的一致性问题的，在逆镜模型中主客观的一致性是不言而喻的。“我觉得最难以理解的事情就是这个世界竟然是可以理解的。”对于爱因斯坦这个难题现在我们可以回答：“因为这个世界就是我们按照易于理解这一原则建立起来的。”易于理解当然符合稳定、经济和可预测原则。

认识论的逆镜模型是建立在“善理”而非“真理”的基础之上，因此感觉的信息加工过程中的“失真”就不再是一个问题。也就是说感觉器官的结构和信息加工的方法并没有“保真”的义务，只负担把信息加工成更好的世界象（或称为世界模型）的任务。分类和识别是制造和使用世界模型的重要步骤，是发生于认识主体中的以善理为根据的过程，所以丑小鸭定理在逆镜模型中不会引起任何理论困难。关于我们的感觉是在求善而不是求真，可以用一个简单的错觉例子加以说明。如图 2.8 所示。图 2.8a 中有三条等距平行线，不论从直观上看还是几何学证明，显然 AG 与 GF 相等。但是把 a 图加上几条线变成 b 图，则看起来 AG 比 GF 要长。这个错觉说明我们的感觉过程对 b 图作了一种立体的解释，把 b 图看成是 c 图折一个角度后形成的三维空间结果在平面上的投影，而对于 c 图来说， AG 比 GF 更长是很显然的。这说明我们的感觉在未被意识之时进行了非常复杂的计算，这是一种必要的能力，没有这种能力我们就不可能从一张照片上看出立体的关系。正是这种能力产生了图 2.8b 的“错觉”。我们的感觉并不“诚实”，但它非常“聪明”。我们的世界模型并不“求真”但是“求善”。这

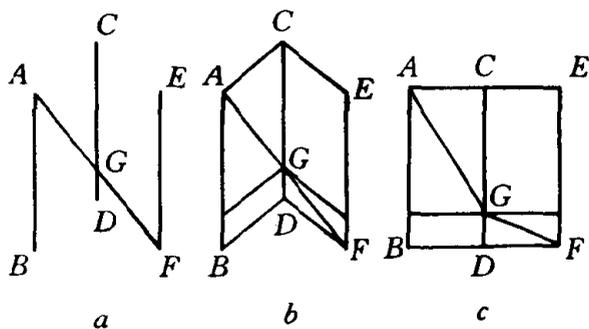


图 2.8

样的例子是不胜枚举的。

刚接触到“逆镜模型”的人可能有不少地方感到难以接受，甚至把它误解为某种唯心主义学说的变种。例如允许笛卡尔妖魔的存在这一点就有可能让一些人感到恐慌。想象我们的生活仅仅是模拟器中的幻象，实在令人不安，假如哪天有人把模拟器的电源关了，那我们不就完蛋了吗？这种担心实属杞人忧天，首先即使假定我们真是生活在笛卡尔妖魔的怀抱里，这个笛卡尔妖魔给我们提供的稳定性也不会小于现实的世界，再说现实世界也并非绝对稳定的，我们都早晚有一天会死，而且不知哪天死，如同佛教所说的“无常”。但是我们也没有理由为此惶惶不安而耽误了享受生活，那样做是非常不经济的。我们可以用爱因斯坦的一句话来形容笛卡尔妖魔：“上帝深奥莫测，但他并无恶意。”逆镜模型只是说没有必要追究笛卡尔妖魔与现实的区别，因为这种区别只对本体论的认识论者才有意义，对非本体论的认识论是没有意义的。非本体论的认识论的一大特点是把主要的注意力放在研究认识主体而不是客观世界，对客观世界只作最低限度的假定。有些人可能认为逆镜模型会导致“唯我论”，即这个世界上只有认识主体自

己存在而他人都是幻影，并不存在。这种想法仍然是起源于本体论的。以我们的认识论的存在观来说，不论是自己还是他人的存在，都只依存于世界模型的稳定性，自我与他人的存在理由是一样的，不如说自我的存在是在理论上更有趣也更困难的问题。以“祭神如神在”的观点看，自我的存在和他人的存在都是认识主体所制造的世界模型中不可缺少的假定，如果不假定自己和他人的存在，那么我写这本书给他人（读者）看就成了没意义的行为，这本书本身也就不存在了。只不过这些存在都没有必要赋予本体论意义而已。

逆镜模型不假定客观世界（认识客体）在本体论意义上的存在，不假定客观世界背后隐藏着规律或理论，不假定主客观之间的先定和谐，但对于认识客体并不是一无所求的。为了讨论在逆镜模型中对认识客体的要求，以及与此密切相关的休谟问题，我们先看两个故事和沃特金斯对归纳的实用问题的解答。第一个故事是大家熟悉的“守株待兔”。农夫正在锄地，一只兔子跑得飞快撞死在田里的树上，农夫不费吹灰之力得到了兔子。于是他不再锄地而每天坐在田里等兔子，兔子再也没来而农夫终于饿死了。第二个故事是“公鸡哲学”。一只小公鸡在进行一个推论：“我在出壳的第一天吃到了米，第二天也吃到了米……第一百天也吃到了米，因此，今天是第一百零一天，我也将吃到米。”养鸡的主妇正在考虑要拿不会下蛋的公鸡做菜，看到这只正在进行哲学沉思的公鸡显得有点发呆，不太活泼，万一是鸡瘟传染起来可不妙，于是一把抓过这只公鸡杀了作成炸八块。第一百零一天公鸡终于没吃着米。这个故事是数学教师讲给学生听的，在数学归纳法之中如果一个定理不能从对于 n 正确推出 $n+1$ 时正确，那么哪怕证明了从 $n=1$ 到 $n=100$ 时都正确，也不能说 $n=101$ 时也正确。

从演绎主义的角度看，“守株待兔”与“公鸡哲学”都是不合乎逻辑的，但是在任何一种依靠实验验证的科学之中，我们都只

能依靠“公鸡哲学”。甚至“守株待兔”的方法也并非不可取，从一个偶然的实验错误引出了大发明在科学史上也不乏其例。这就是归纳的实用问题，即我们在决定未来的行动方案时以过去验证成功的理论为指导为什么是合理的？

以善理为基础的逆镜模型由于不包含真理概念而避免了理论上的休谟问题，不会发生一方面说要追求真理而一方面又证明不可能得到真理的令人沮丧的结局。但是归纳的实用问题依然存在，即我们有什么理由期待过去表现很好的理论在将来也表现好，过去预测成功的理论在将来仍能预测成功。从逻辑上看，这仍然是一种公鸡哲学。沃特金斯对此作了非常巧妙的论证，他利用了在不确定条件下作出合理抉择的一个假定，即在不确定情况，选择对未来更弱的预测为前提的方针是合理的。因为如果其他方面相等，一个更弱的预测比一个更强的预测的错误可能性更小。按照这个假定，如果我们认为过去验证成功的理论在未来会不成功，那么我们就等于作出了一个很强的预测，即客观世界在现在发生了重大的变化，而认为过去成功的理论在未来仍然成功就不需要作这个预测。但是这个论证仍然使我们感到是在用公鸡哲学证明公鸡哲学。因为我们认为客观世界现在发生重大变化的预测与不发生重大变化的预测相比是一个更强的预测这一点本身只能从公鸡哲学中找根据。

在逆镜模型之中，我们用对认识客体的基本假定来论证公鸡哲学的合理性。对于逆镜模型的认识客体所作的要求比其他认识论要少得多，但是仍然有一点最低限度的要求，即认识客体必须向认识主体提供相对稳定的相关关系。从空间上说，客观世界向认识主体提供的输入信息与认识主体的输出信息不能是完全无关的。从时间上说，认识主体先后得到的输入信息之间不能是完全无关的。所有这些关系应具有相对的稳定性。公鸡哲学的合理性就建立在这种稳定性的基础之上，但是由于这种稳定性是相对的

而不是绝对的，所以公鸡哲学在逻辑上不能保证我们绝对成功。不过公鸡如果用第 5 天之前的经验去预测第 6 天之后的事，那么还是有 95 天是成功的。

我们怎样才能证明对客观世界所作的稳定相关假定是成立的呢？可以用归谬法（反证法）来证明。假如一个认识主体面对一个非常糟的笛卡尔妖魔，这个妖魔毫不理睬认识主体的输出信息，对认知主体的输入不发出任何信息或是只发出白噪声（对白噪声不熟悉的读者可以在电视台不播放的时间打开电视机，开大音量，就可以看到和听到近乎白噪声的东西。）那么这个认识主体就不可能建立起任何一个稳定的世界模型。这个认识主体能不能认识到“这个世界是个白噪声源”呢？不可能。因为这个认识主体连“白噪声”的概念也建立不起来。从进化的角度看，这个世界如果不具备相对的稳定性，进化中被选择的一切性状在其后都会立即失去优越性，一切生物都会灭绝，当然人类也不会产生。但是这种稳定性不是绝对的，所以在生物进化史上灭绝了的物种比现存的要多得多。这也是公鸡哲学的宿命。人类作为认识主体本能地厌恶白噪声，由此可以理解为什么爱因斯坦不喜欢让上帝掷骰子。

对客观世界的稳定相关假定与认为客观世界是有规律的有什么不同呢？其不同点在于，如果从本体论意义上考虑客观世界所固有的规律，那么这个规律就具有唯一性和绝对性，认识的任务就是发现这些规律；而稳定相关假定只提供认识主体以建立世界模型的原材料，从这些材料出发所能建立的模型是有无限多种的，理论的产生不是靠发现，而是靠发明。这种说法有时会引起一种误解，好象认识在很大程度上是凭空捏造的，那么就很难解释为什么很多人会对客观世界有大致相同的看法。对于这种共性可以有多个方面的解释：首先是人类生活在共同的相对稳定的物理世界和文化环境之中，而人类作为一个生物种具有大致相同的基因组合，因而有大致一致的由基因决定的认知基本结构，而且具有共

同的生物价值观（三大原则），再加上人们绝大多数的认识来自于对文化传统的继承，对于个人来说自己发明的只是极少部分。这就是人们可以取得共识的原因。但是正如每个人都有不同的相貌一样，每个人的世界象都会由于基因、经历、局部文化环境、文化层次的价值观的不同而有所不同。

在本章的开头我们提出了建立工程认识论的任务，即要通过本章建立起具有工程可实现性的认识论。为此我们首先通过比较文化学来建立非本体论的认识论概念，然后说明从认识论中铲除形而上学的必要性，并建立善理的理论基础三大原则以代替真理这一形而上学的集中表现，最后提出逆镜模型作为工程认识论的具体形式。工程认识论作为非本体论的认识论，把本体论所引起的形而上学问题转化为认识论中的非形而上学问题，为认识和认知的研究确定了科学的方向。例如我们不再问“月亮在我们不看它的时候是否存在？”而是问“为什么在我们不看月亮时也会认为它存在？”，我们不再问“概念或符号与客观事物具有怎样的对应关系？”，而是问“在认知过程中概念是怎样产生的？”，我们不再问“机器能有意识、感觉和感情吗？”而是要问“意识、感觉和感情在认知过程中起什么作用？怎样在机器中实现这种作用？”，我们不再问“机器能有自由意志吗？”，而是问“人为什么会有自由意志的错觉？如何在机器中实现这种错觉？”等等。这种排斥形而上学的科学态度将作为本书全书的指导思想，在后几章中问题的设定都将依据这种态度。

总而言之，工程认识论把“怎样才能认识客观世界的本来面目？”的问题变为“认识主体怎样建立起客观世界的有效模型？”的问题。讨论前者可能会引起诸多形而上学的、科学所无能为力的问题，而后一个问题则没有这个危险。当然，后一个问题的工程可实现性还不够明显，排斥形而上学仅仅是工程可实现性的必要

条件，并不是充分条件。我们所提出的输出与输入信息的稳定相关条件只构成一个开端，而存在原则、经济原则和预测原则只是一组约束条件，由此出发到达建立世界模型还有很多未知的部分，但这不再属于哲学层次，将在后几章作为认知科学的问题进行探讨，从而对于意识、感情、感觉、分类、识别、自由意志、创造、学习、灵感、欲望、思考、熟练等认知过程中的各种现象给出机械论的解释。但是哲学层次的结果也已经能够使我们对于认知科学中某些研究方向所遇到的困难作出解释，例如符号主义和逻辑主义，其符号与现实之间的对应关系问题是一个哲学上的形而上学问题，不可能用科学的方法解决，这是无法制造常识系统的根本原因。符号与逻辑是认知的产物而不是认知的根据，因此符号主义路线不可能引导我们走向对认知过程的深入理解。

近代物理学的发展，相对论和量子力学的建立，使得“观察者”（认识主体）成为物理学中不可忽视的一个因素，由此也引起了一些困难的哲学问题。这些困难是不是由本体论以及建立在本体论基础上的认识论所引起的？我们在此提出的认识论观点是否可以对“薛定谔的猫”作出合理的说明？这是个令人感兴趣的问题。但是由于这个问题与本书目标关系不大，（更诚实地说是由于笔者对量子力学知之甚少）在此不能作深入探讨，只留下一个“接口”，以引起对此有兴趣的人进行思考。

第三章 广义进化论

3.1 自我表述系统

对于人工智能的研究，曾经有过一个颇具讽刺意味的评论：“为了登上月球的目的，爬到树顶上，就说是前进了一步。”这个比喻听起来很可笑，但是我们不得不承认它是非常贴切，入木三分的。拿“专家系统”来说，专家系统无疑在一些领域有其实用价值，但是用专家系统显然难以解释人的认知过程，专家系统利用开发者放到系统中的知识来解决问题，但却不能象人那样在解决问题的过程中增加知识。这也就意味着专家系统的能力不可能超过“树”也就是系统开发者放到系统中的形式化知识的范围。在人工智能领域中曾经有过制造“通用问题求解系统”的目标，但是人除了解决问题之外，还具有提出问题的能力。因此我们不应仅仅关心“通用问题求解系统”，还应该关心“通用问题产生系统”。为了深入理解人的认知过程，注意力不应放在“如何爬树”的问题上，而应当去研究“树”本身是如何长起来的。爬树爬得再高，也爬不到月亮上去，但是如果我们能揪着自己的头发哪怕能离地一寸，就有可能把自己提到月亮上去。用更符合物理法则的比喻来说，我们应该研究象火箭那样的不依靠其他支撑而能够自行升空的装置，以便增进对于人类这样的认知主体所具有的创造性的认识。在工程上有个与此相对应的概念叫作“自举”，例如在电子线路中有“自举电路”。计算机的起动过程也是一种自举过程，电源接通后先由只读存储器中的一段小程序把外存中的指定部分读进内存，然后运行这部分读入的程序再读入操作系统的其

他部分，直至操作系统自己把自己安排就绪，进入等待用户指令的状态。工程中的自举通常并不包含什么创造性，但是它们有一个共同的特征，就是都含有某种环状的结构。自举电路可以看成是一个正反馈环，计算机的起动过程中，被读入的程序可以读别的程序，也包括自己。“自举”的概念可以说是“自我表述”的一个初级形式。

以“自我表述”为基准，可以把科学所研究的对象分为两类，一类是“自我表述”的对象，一类是“非自我表述”的对象。物理学是非自我表述对象的典型例子，牛顿力学的建立并不改变行星的运行轨道。社会则具有自我表述的明显特征，社会科学的研究成果，总会对社会本身的发展产生或多或少的影响，“孔子作春秋，而乱臣贼子惧”。人脑的认知过程也是一个典型的自我表述系统，人的每一次思考的结果都会影响今后的思维方式。与“非自我表述”的对象相比，自我表述的对象研究起来要困难得多。首先是由于自我表述的对象变化多端，稳定性很差，科学所要求的可重复性经常难以满足。对于非自我表述的对象，我们可以先建立浅层的稳定认识，然后以此为根据去建立深层的更加稳定的认识，逐步深化。但是对于自我表述的系统，浅层的认识难以稳定，深层的认识也就失去了根据，因此很难对于社会科学建立起象牛顿力学那样的深刻的理论。正因为这个原因，我们至今缺乏一种针对自我表述系统的研究范式，如同在非自我表述领域中用欧几里德范式建立牛顿力学那样。

本章的目的就是为研究自我表述系统提供一个研究范式，或者说提供一个通用的模型，以便研究富有创造性的过程。这类富有创造性的过程包括社会结构的进化、文化的进化、人脑的认知过程、生物的进化、生物个体发育过程等等。在这些领域之中，我们选择生物的进化作为建立自我表述系统研究范式的基地。这是因为生物进化是自然界中最富有创造性的领域；作为自然科学，可

以天然地避免形而上学的入侵；而且从达尔文以来，进化生物学在长期的研究之中积累了极其丰富的研究成果，对于建立理论来说提供了良好的基础；达尔文的进化论本身也已经是一个相当深刻的理论，因此一百多年来关于进化生物学的研究能够不断取得硕果而并未脱离达尔文的基本思路。

但是对于我们的目的即建立自我表述系统的研究范式来说，直接搬用达尔文的进化论是不够的，因为在达尔文的进化论理论之中并不包含自我表述的概念，为此本章的主要任务就是在达尔文进化论的基础之上加入自我表述的模型，以建立一种包含自我表述概念的进化理论——广义进化论。在第一章我们曾经用打麻将来比喻进化的过程，把这个比喻推广一下，广义进化论就相当于一种更复杂的麻将，在这种麻将之中，每一个牌局结束时，赢家牌的组合将会以某种方式影响下一局打牌的规则，这会导致一种极其复杂的自我表述过程。显然，这种过于复杂的游戏作为娱乐工具是难以被接受的，而且作为自我表述的模型也因太复杂而难以理解，因此我们仅仅把它作为一种比喻，以说明自我表述系统与非自我表述系统在复杂性方面的区别。

从系统论的角度来看，引进自我表述的概念是为了研究一种具有高度复杂性的系统。通常对于系统的复杂性，我们容易看到系统的元素数量大，以及元素之间的关系的多重性。例如网状的关系比起树状或直线状关系就要复杂得多。但是有些元素数量巨大的网状相关的系统有时表现并不太复杂，例如一个世界规模的电话网，其元素（电话用户）的数量可达几十亿，接近人脑神经元的数量，其互联方式除了用户与电话局之间是树状之外，电话局之间也是网状连接，但是电话网并不使我们感到太大的复杂性，因为电话网在工作过程中并不改变自己的基本结构和工作原理。包含自我表述的系统，其表现要复杂的多。这种复杂性最突出体现在系统在工作过程之中自己能改变自身的结构甚至工作原理，

为了把握这种系统，我们就需要一个更为深层的，在表层发生变化时保持一致的模型。这就是我们将要在这一章中提出的软硬结构模型。软硬结构模型的基本运行机制是进化论的变异与选择。按照达尔文的进化观点，变异是随机产生的，可以在系统中引进随机因素而简单地实现。而选择的问题就要复杂得多，选择必须依赖于某种价值观或价值体系，建立评价准则是掌握进化过程的关键，所以在这一类复杂系统中，评价是最核心的问题，这也就是我们在哲学层次中要建立以评价为中心的善理体系的原因。把软硬结构模型和自我表述概念引进达尔文的进化论，就构成了广义进化论。广义进化论是对达尔文进化论的推广，而其新加入的软硬结构模型又依存于进化的工作原理。进化依赖于评价，但是评价问题在达尔文进化论中并不突出，这是由于在生物进化过程中评价问题比较简单，最根本的评价准则就是存在原则。“优胜劣败，适者生存”从逻辑学角度看只是循环论证而已，说胜者为优而优者必胜是没有太大意义的。这句话从善理的角度看则表现了一个最基本的评价准则，即存在原则。从这个原则出发，就可以对生物的适应度给出严密的定义，即某种生物的适应度是指该生物的一个个体在一生中能够繁殖出的达到再生育状态的子个体的个数。用数学语言来说是生物个体指数增长时指数函数的底，即 $M=a^N$ 中的 a ，而 N 是繁殖多少代。因此虽然在达尔文进化论中评价问题不很复杂，但进化建立在评价准则的基础之上这一点是显而易见的。建立广义进化论的目的在于研究自我表述系统，对于某些自我表述系统来说，虽然存在原则仍是其评价的最基本原则，但更直接起作用的却往往是一些派生出来的次级准则，或称为高层准则。例如当我们用广义进化论来分析认知过程中的创造和学习的机制时，存在原则是不能直接起作用的，死亡是进化中的选择手段，却不能成为创造和学习中的选择手段，生物的学习是生存期间的过程，生物不可能通过自身的死亡来学习任何东

西。而次级的评价准则本身又可能是进化与学习的产物，因此关于复杂评价体系的研究的重要性在广义进化论中要比在达尔文进化论中更为突出。

“自我表述”这个概念来自于道格拉斯·霍夫斯塔特 (Douglas R. Hofstadter) 所著的《GEB——一条永恒的金带》。该书的中文摘译本中多用“自相缠绕”、“自我相关”、“怪圈”来表达这一概念。为了更深的了解这一概念，建议读者去读一读原书。在本章中提出的软硬结构模型可以说是“怪圈”概念的一个具体化的工程实现方法。“怪圈”这个词相当好的表现了这个概念，遗憾的是这个词汇已经相当多地出现于中国的大众传媒之中，而代表着类似于“恶性循环”或“死锁”这样的其他颇具贬意的概念。因此我们在此不得不放弃使用“怪圈”这个词，以避免引起误解，代之以“自我表述”这样一个还不太为人所熟悉的词汇，使读者从不带偏见的状态出发，逐步建立正确的概念。

3.2 进化的地形图模型

在讨论广义进化论之前，我们先介绍一些在经典的达尔文进化论研究中经常使用的模型和方法。由于广义进化论包含达尔文进化论，这些方法对广义进化论也是不可缺少的。本节介绍地形图模型。地形图模型在数学领域通常与最佳化相对应，用这个模型来表现进化过程也意味着我们把进化看成一个最佳化的过程。地形图模型最简单的形式示于图 3.1。图中所示的是一个一维的地形图，这个图也可以看成是多维地形图的一个断面。二维地形图就是通常我们所看到的地图的形式，一般用等高线来标出高度，三维或更高维的地形图在纸面上表现就比较困难了，可以通过数学的方法来掌握，或是表现成很多断面的形式。在一维地形图中有三种地形，图 3.1 中 A 点的“坡”、B 点和 D 点的“坑”和 C 点

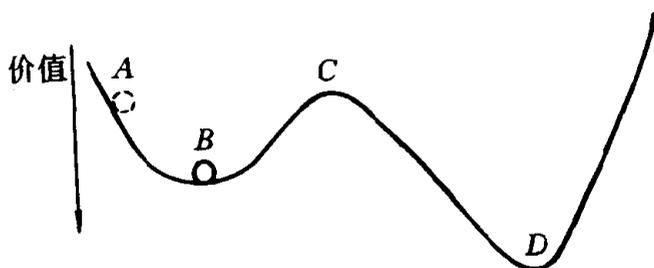


图 3.1

的“峰”。在二维和多维的地形图中还有在某些方向是“坑”而在另一些方向是“峰”的马鞍形的点，可以称之为“山口”，从一个坑走到另一个坑时，通过山口是比较省力的。在图 3.1 的地形图中处于 A 点的坡上的球受重力的作用滚到 B 点的坑里，就表现了一个最佳化的过程。在这个过程中，重力代表了一个评价准则，即位于低处的价值大于位于高处的价值，也就是说，表示价值升高的箭头方向是指向下方的，与重力的方向一致。这似乎与我们的习惯不太符合，似乎应该把价值的箭头指向上方才对。对于一个模型来说，坐标设定的翻转对于模型的表现能力没什么大影响，因此箭头向上还是向下原则上没什么不同。在数学中的求最佳值的算法有时称为爬山法，有时叫作最快下降法，说明这两种坐标方向设定都被人采用。中国有句俗语：“人往高处走，水往低处流”。在此我们宁愿选择水这种比人简单得多东西作为模型的介质。这反映出我们一贯的反形而上学的哲学态度，或者说在形而上学家看来是“堕落”的哲学。也反映了一种用低层次现象去解释高层次现象的科学中普遍存在的还原论倾向。从实用上看，“堕落”模型更容意防止一种“这山看着那山高”的误解。当我们爬上一座山峰，就可以看见别的更高的山峰，但在最佳化计算之中更高的

山峰是看不见的，不能忘记“爬山法”的全称是“瞎子爬山法”。瞎子的“明杖”只能探知近处的高低，稍远一点的山都是看不见的。下降的模型就不易产生这种误解，因为在一个坑里通常发现不了别处还有更深的坑。这体现了最佳化算法的难处，即所谓的“局部极小点”。在坑不止一个的地形里，我们常常会掉在一个浅坑里而找不到最深的坑。

进化的地形图模型可以解释进化中的很多问题，例如进化中的急变与缓变。化石研究为进化提供有力的证据，也提出了很多问题，其中一个问题就是化石所描绘的进化过程不是平缓的渐变过程，而是稳定与跳跃的交替。我们往往能找到某些物种的大量化石，而从一种物种到另一种物种的中间状态化石证据却很难找到。这就为神创论准备了一个藏身之处，即物种并非由进化产生，而是由上帝分多次创造的。用地形图模型可以简单地说明跳跃与稳定的过程。某个物种的长期稳定存在是这个物种位于坑里的状态，由于某种因素（我们可以想象地形图受到大的震动）某个球偶然落到空坑旁边的坡上，球当然会很快地掉到这个空坑里，在里面繁殖起来。即使地形图经常受到或大或小的震动，我们观察到球在坡上的机会仍然很小，这种中间状态的数量少而且时间短促，因此象始祖鸟这样的化石是非常罕见的。总鳍鱼这样的坡上物种能活到今天，可以说是掉在坡上的一个小坑里了。因此我们看到的物种或化石之间并不显示出连续变化的迹象。接受神创论还是接受进化论似乎只能靠信仰，无神论也无非是一种形而上学。从真理的角度看神创论和进化论都不会被确证为真理。因此不久前在美国仍有人提出神创论也是一种理论，应该在学校里与进化论一样作为一种学说来讲授。但是从善理的角度看，神创论显然算不上一个合格的理论。进化论可以回答物种怎样产生的问题，而神创论对于上帝怎样造出物种这一点一无所知。如果我们问豆腐从何而来，我们所关心的是豆腐的制造工艺过程，豆腐的发明者

的姓名对我们并没有太大的意义。

更细致地分析地形图中的进化过程，用球滚下坑的比喻就显得不够严密了。对于一种由多基因决定的具有连续性的性状（例如人的身高）的进化，可以用在坡上的液滴来形容。对于连续性性状可以用正态分布来近似，正态分布曲线的形状象一个浸润状态的液滴。由于液滴处在坡上，各方向的适应度不同，下一代在适应度较高的方向（坡的低处）个体频度增加，导致正态分布的平均值向该方向移动，这样经过多代之后，液滴流到坑底，进入稳定状态。现存的绝大部分生物性状都可以认为是处于坑底的稳定状态，缓慢的进化往往是地形缓变导致坑底的缓慢位移造成的。

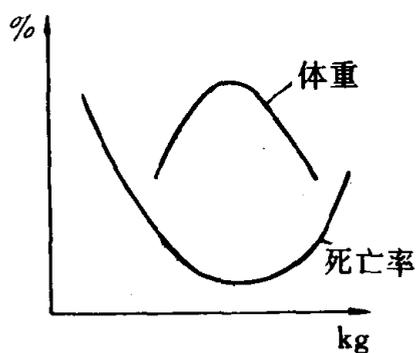


图 3.2

例如在图 3.2 中表示的新生儿体重分布曲线和死亡率与体重的关系曲线。从图上可以看出新生儿体重分布的平均值接近于死亡率的极小点，处于稳定状态。新生儿体重越小死亡率越高是容易理解的，可以假定是由于虚弱造成的。体重过大死亡率升高如果假定是由于难产的增加所引起的，那么随着剖

腹产技术的推广，大体重新生儿的死亡率下降，死亡率曲线的极小点向右移动，新生儿体重分布也随之向大体重进化。一个简单化的推论就是，几万年之后，人将进化到离了剖腹产就不能生孩子的地步。当然这只是个模型分析的例子，不能由此得出应该禁止剖腹产的结论。因为难产增加未必是大体重新生儿死亡率高的唯一因素，医疗条件的改善同时也降低小体重新生儿的死亡率，作过剖腹产的妇女不宜多次生育也足以抵消新生儿死亡率下降的影

响，而且由于新生儿死亡率本身非常低，不可能成为造成强进化压力的因素。所以上述简单化推论是不能成立的，正向和反向的“九斤老太”的看法都站不住脚。

地形图模型在进化论中非常有用，但是也有些问题用地形图模型难以解决，有时甚至造成误解。如果用震动地形图模型来解释从古至今的全部进化过程，就很容易给人造成一种极其错误的印象，即从古至今的进化过程都是事先安排好的。坑原来就在那儿，只是由于震动的随机性，使得能繁殖的小球或早或晚掉到坑里去，最后占据所有的坑。这是一种彻底的预成论或是宿命论的想法，似乎进化只是卷着的地毯展开的过程。而实际的进化过程要比这种简单想法复杂得多，是充满创造性的过程。也有人提出过动态的地形图模型，把地形图想象得象水面的波浪那样向前涌动，推着物种的小球前进。但是这个模型显然不太好用，描述一个不断变化的地形很困难，而且在这个模型中不包含变化的原因，顶多只能考虑外因，对于进化自身的创造性仍难以容纳。在实际的进化过程中，进化的结果本身会改变地形，而且不只改变地形本身，还会改变地形图的维数，即增加新的评价对象因素，或取消某些已经不重要的因素。如果考虑这些因素，地形图模型将变得复杂不堪，难以使用。因此地形图模型最适用于研究少数因素的一次性进化过程，或者说是少量参数局部寻优的过程。遗传算法就是基于地形图模型的一种最佳化算法，其中采用了突然变异和模拟有性生殖的基因交换等进化方式，但是这一算法与其他算法相比并没有显示明显的优越性，突变不比爬山快，基因交换也未必强于随机搜索，这种现象反映了地形图模型本身的局限性，也反映了对于有性生殖的意义并未理解。生物的进化过程本身改变地形图的例子比比皆是，例如由于生命活动把原始的还原性大气变成氧化性大气，以及共进化过程，甚至生存竞争本身就造成地形图的变化。

3.3 进化论的博弈论方法

在进化生物学中引进博弈论方法是进化论研究从 70 年代以来的一个重大发展。本节介绍博弈论方法主要是依据梅纳德·史密斯所著的《进化与博弈论》一书(John Maynard Smith, *Evolution and the Theory of Games*, 1982, Cambridge University Press)。本节不可能详细介绍这一理论,特别略去了严密的数学部分,有兴趣者最好参考原著。博弈论与生物学的结合,不仅促进了进化生物学的发展,而且为博弈论找到了最佳的用武之地。早期博弈论被应用到政治、经济等领域,用以研究人的合理决策过程,但是并未取得预期的成功,主要是由于以下两个原因:第一,博弈论要求一维的价值观,而对于人来说要把多维的价值观(经济利益、死亡的危险性、避免良心谴责的喜悦等等)统一到一维时,很难避免不自然的任意性。而在生物进化之中,达尔文的适应度可以成为名符其实的一维评价尺度。第二,也是更重要的原因是,人总是按照最合理的方案行动这个假设前提是很可疑的,但在生物学中用进化的稳定性概念取代了人的合理性,而进化使生物集团处于稳定状态这一点具有充足的理论根据。因此博弈论在生物学中的应用比分析人的行动更容易成功。

对于本书所关心的问题来说,博弈论的意义超越了梅纳德·史密斯的预想。从地形图模型的角度看,博弈论是一种处理变化地形图的方法。对于我们最关心的建立价值观体系的要求来说,博弈论提供了一种具有严密数学背景的求取评价函数的重要方法。博弈论用于生物进化时最重要的概念是“进化稳定战略”(Evolutionarily stable strategy, ESS),梅纳德·史密斯并不倾向于把稳定与某种价值观联系起来,他认为“平衡点并不对物种的延续是否有利作出判断”,但是我们非常重视 ESS 概念在价

价值观体系中的作用，恰巧我们的价值观体系正是建立在存在（物种延续）和稳定（平衡点）概念的基础之上的。*ESS* 概念具有普遍性，很容易推广。道金斯（Dawkins, 1980）强调过同样的想法可以导出“发展稳定战略”（*developmentally stable strategy, DSS*）以及“文化稳定战略”（*Culturally stable strategy, CSS*）等概念。而且 *ESS* 概念还可以推广到学习领域，找出“进化稳定学习规则”这种很通用的规则。无论何种学习规则，都需要以某种评价为前题，如果评价准则错了，就会引导学习走向不适应的方向，从而使带有这种评价准则的生物个体被淘汰，因此价值观体系可以由进化和文化进化的过程建立起来，而无需外在的形而上学基础。研究进化的地形图模型和博弈论方法可以推广到价值观的建立过程中来，只要用选择行为方式的评价准则来代替表示具体行为方式的战略即可。用这种观点研究价值观体系，可以把 *ESS* 和 *CSS* 的概念推广为“进化稳定价值”（*Evolutionarily Stable Value, ESV*）和“文化稳定价值”（*Culturally stable Value, CSV*）的概念，为价值观体系的研究找到具有严密数学背景的方法。进化稳定价值（*ESV*）是文化稳定价值（*CSV*）的基础，生物进化比文化进化要缓慢和稳定得多，文化的价值观一般说来应与生物价值观一致，有时表面看起来这两种价值观之间会发生冲突，但是从长远或全局来看，文化价值观或者与生物价值观一致，或者逐步被淘汰，虽然淘汰过程可能很长。*ESV* 和 *CSV* 的概念为我们提供了一种可能性，即通过 *ESV* 和 *CSV* 用博弈论在生物学中适用的手法建立起有科学根据的价值观体系，从而为博弈论用于象人类社会这样的拥有复杂价值观体系的系统开辟道路。

梅纳德·史密斯对进化稳定战略所下的定义是：*ESS* 是指如果集团全体采用该战略，则由于突然变异而产生的采用其他任何战略的个体侵入这个集团，都会被自然选择所排除。稳定通常被看作一种绝对性的概念，一个系统或者稳定，或者不稳定，很难

处于中间状态。但是当我们把稳定看作善理的基础，由于善恶好坏是相对的概念，因此就必须重视进化稳定战略这个概念的相对性。我们可以从以下两个方面来说明这种相对性，第一，虽然在定义中是说“其他任何战略”，但是在实际的分析过程中我们只能建立包含有限个战略的数学模型，不可能包括将来所有可能产生的战略，也就是说我们可以在有限种战略中选择最好的战略，却不易证明将来不可能出现更好的战略。第二，即使我们可以用数学的方法严密地证明某种战略为进化稳定战略，也要注意这种证明是在一系列前题条件下才能成立的，而进化是一个充满创造性的过程，前题条件是很容易变化的。

以下我们通过具体例题来简单介绍进化中的博弈论方法。第一个例题是“鹰鸽博弈”。假定两只动物为争夺具有价值 V 的资源进行战斗。此处的价值 V 指的是动物个体得到这一资源时能使其达尔文适应度增加 V 。注意当战斗失败而未获得资源时其适应度也未必为 0，例如资源可以是一个好的生活领地，在好的领地上繁殖平均存活子代数为 5，在坏领地上子代数为 3，此时 V 就是 $5 - 3 = 2$ 。假定在战斗中动物可以有三种行为方式：“恫吓”、“战斗”和“逃跑”。战斗意味着使对方受到伤害，而逃跑则是把资源让给对方。现实中动物争斗时战略可以很复杂，我们先假定动物只能采取以下的两个战略之一，而且一个个体总是用相同战略争斗。

“鹰战略”：坚决战斗到对方逃走或受伤。

“鸽战略”：先恫吓，如对方战斗则逃走。

当双方都坚决战斗时会造成一方负伤逃走，此时假定负伤使适应度下降 C 。用 H 代表鹰战略，用 D 代表鸽战略，可以写出如表 3.1 所示的利益矩阵 (*payoff matrix*，通常译为“支付矩阵”，为表示正值代表获益，称之为“利益矩阵”)，其中表示了争斗所引起的适应度增减。

表的左侧表示自己所采取的战略，上方则表示对方采取的战

略。表中取值的理由是，当鹰派与鹰派争斗时，各自获得资源和负伤的概率为 0.5，鸽派与鸽派争斗则各自获得资源的概率为 0.5，鹰派与鸽派争斗则鹰派得到资源而鸽派逃走。

表 3.1

	<i>H</i>	<i>D</i>
<i>H</i>	$\frac{1}{2} (V-C)$	<i>V</i>
<i>D</i>	<i>O</i>	$\frac{V}{2}$

一般性的分析结果表明，象鹰鸽博弈这种有两个单纯战略的博弈必然存在 *ESS*，利益矩阵如表 3.2 所示时，如果 $a > c$ 则 *H* 是 *ESS*，

表 3.2

	<i>H</i>	<i>D</i>
<i>H</i>	<i>a</i>	<i>b</i>
<i>D</i>	<i>c</i>	<i>d</i>

如果 $d > b$ 则 *D* 是 *ESS*，当两个不等式同时成立时，则 *H* 和 *D* 都是 *ESS*。两者都为 *ESS* 是这样一种状态，即如果集团都由鹰派组成，则少数的鸽派难以侵入，而少数鹰派也不能侵入鸽派集团。当两个不等式都不成立时，即 $a < c$ 且 $d < b$ 时，以 *P* 的概率采用 *H* 战略，而以 $1-P$ 的概率采用 *D* 战略的混合战略是 *ESS*。

$$P = \frac{(b-d)}{(b-d) + (c-a)}$$

把这个结果用于表 3.1 所示的具体情况可知，由于 $V > V/2$ ，所以鸽战略不可能是 *ESS*，软弱者遇到强者总是吃亏。当 $V > C$ 时，*H* 是 *ESS*，即如果资源的价值大于争斗成本时，争斗总是上

算的。当 $V < C$ 时，混合战略是 *ESS*，混合战略既可以是每个个体都以某种概率分别采用两种战略，也可以是在集团中一定比例的个体只采用 *H* 战略而另一部分只采用 *D* 战略。这种 *ESS* 意味着当采用 *H* 战略的总概率高于平衡点时，*D* 战略上算，即多数强者都被争斗弄得两败俱伤时，弱者可以坐收渔利；反之弱者多了，强者争斗损失变小，强者上算。在资源价值 V 不变时，争斗损失 C 越大，则对鸽派越有利。当动物持有一招致命的武器时，在种内争夺之中往往不使用，例如一种剧毒的蛇可以一口把另一条蛇咬死，但是在争夺配偶的争斗中从不使用毒牙，而是两蛇并排挤靠角力，力小者甘拜下风。即使由于突然变异可能产生动用毒牙的个体，争斗时的巨大损失会自动地把这种基因从种群中淘汰掉。这也可以说明为什么武器的杀伤力超过某种限度时，反而不能用了。在 $V > C$ 的条件下，*H* 是 *ESS*，种群全由鹰派组成，如果考虑个体之间战斗能力的差异，自然选择必然促进军备竞赛，使得争斗成本 C 增加，最后使不等式倒转，导致鸽派的侵入，进入混合战略，因此 $C > V$ 是更为常见的状态，即鹰派与鸽派通常是共存的。

以上简单的对称博弈增加一些条件就可以构成非对称博弈。例如假设某个领地对于持有者和入侵者具有不同的价值，持有者由于熟悉领地情况，领地价值较高，如果失败而被逐出，在新的领地上要花费精力重新了解情况，损失较大，而侵入者即使失败，改换别的领地也和现在争夺的领地一样是未知的，因而领地对于持有者的价值大于对于入侵者的价值。再假定动物能够区分自己的立场。这样在鹰战略 *H* 和鸽战略 *D* 之外再加上布尔乔亚战略 *B*。布尔乔亚战略是当个体在争斗中处于持有者立场时采取鹰战略，而在处于入侵者立场时采取鸽战略。可以证明，在这种三战略博弈中，布尔乔亚战略是 *ESS*。布尔乔亚战略可以解释私有现象的产生，如果一个原始人捡到一块石片，经过长期的摸索掌握了石片的各种使用方法，那么这块石片对于持有者就比对于其他

原始人有更大的价值。持有者会不肯把石片交给他人，而他人也会承认这种占有。反之，随着社会的发展，技术变得日趋复杂，一个人只可能掌握少数的技术，而掌握这种技术之后制造某一产品要比别人花的劳动少，别人使用这种产品又很容易，产品的交换就成为一种有价值的行为了。如果生产任一种产品所有人所花费的劳动都相同，那么赶集就是一种徒劳无益的活动。交换的价值也就造成了分工的价值。

另一种非对称博弈是假定动物能够在争斗之前对双方力量进行评估，在某种程度上预测争斗的结果。例如动物可以通过看对方的体形大小来获得关于战斗能力的信息。此时博弈可以在原有的鹰战略和鸽战略之外再加上评估者战略 *A*。评估者战略是在争斗之前评估双方的战斗力，敌强我弱时采取鸽战略，敌弱我强则采取鹰战略。这是一种典型的欺软怕硬战略。可以证明，在 *H*、*D*、*A* 三种战略的博弈中，评估者战略是进化稳定战略。由于评估者战略是按照实力大小分别使用鹰鸽两种战略，所以评估者在情绪上必然是恐惧与无畏并存，在评估过程中根据对方的情况两种情绪的一方压倒另一方。由评估者战略是 *ESS* 这一点可以很好地解释动物在斗争中的仪式化倾向。在六十年代，生物学界的主流观点认为仪式化是由种进化和群选择造成的，仪式化可以减少种内斗争的损失而提高种群的适应度。现在则认为这种解释不对，在生物的进化中有很多降低种群适应度的性状，只要能提高基因适应度，都能成为 *ESS*，扩散到整个种群中。这就是“自私的基因”学说，生物进化过程中产生的绝大部分性状是“自私”的，即缺乏远见和全局观念的。当然种群适应度并非不起作用，但要在很长的时间尺度上才看得出来，而且是通过基因适应度间接起作用的。动物在斗争之前的“仪式”是一个相互评估的过程。动物在仪式之中尽量显示决定战斗力的关键因素，例如角、犬齿、体形较大的一面，即所谓“张牙舞爪”。有的动物在发怒（发怒是一种临战

状态)时毛会竖起来,是由于竖起毛来会使自己看起来体形更大,增加敌方评估中的恐惧因素。人类由于有毛的部位少,竖毛作用不大,此种功能已经退化,但是仍有少数人保留了这种性状,所以才会有“怒发冲冠”之说。动物在临战状态往往发出低沉的吼声,而在战败或负伤逃走时往往发出尖锐的悲鸣,这是由于大的物体发声频率低,低沉的声音可以造成体形较大的评估,而高频率则是示弱,告诉对方胜负已分,穷寇勿追,以减少双方争斗成本。因此在音乐中重低音给人以力量的感受。

像布尔乔亚战略和评估者战略这样的非平衡博弈,要求参战者具有正确评价资源价值或双方战斗力的能力,这也就是在动物中评价能力得以提高的进化压力。高等动物的复杂的评价体系就是在这种进化压力下逐步形成的。动物通常不识数,但对于量的判断能力很强,因资源价值和战斗力往往与某种物理量相关。

第二个例题是个经典的问题,叫作“囚犯困境”(有时译为“囚徒悖论”或“囚徒疑难”)。问题是这样的:设两个合伙犯罪的人被隔离审讯。他们知道,如果他们都不招供,警方没有足够的证据判他们重罪,只能判2年;如果他们都招供了,均将被判9年。然而,若一个人招而另一个人不招的话,招者将获释,不招者将被判10年。在隔离的情况下,这两名囚犯应该怎样做才对自己有利?其中一名囚犯可能这样想:“如果对方不招,我招了被释放,我不招被判2年,我招了有利;如果对方招了,我招供被判9年,不招是10年,还是招了对我有利。”于是这个囚犯决定招供,另一名囚犯出于同样理由,当然也招供,结果是两人都被判刑9年。但是如果两人都不招供,显然是一种更好的结局,两人都判2年。如果套用表3.2的一般性结果,结论也是一样,招供是唯一的进化稳定战略。这个经典问题是在1950年左右由兰德公司的梅瑞尔·弗拉德和梅尔温·德雷舍(Merrill Flood, Melvin Dresher)发现的。这是个困难问题,但又

是日常生活中常常遇到的问题。例如在公共汽车站上车少人多时的“乱挤”现象，就相当于多人的囚犯困境博弈，如果大家都排队，即使这一趟车上不去，下一趟车可以先上，这是全局最佳战略，但不是进化稳定战略，只要有少数乱挤者就可以破坏整个秩序，形成全员一拥而上的局势，结果所有人都多花了力气和时间，而不乱挤的人过好几趟车也上不去。当然这个例子比较复杂，其中还包括了鹰鸽博弈、评估者战略和军备竞赛等多种因素。囚犯困境的数学表现如表 3.3 所示。这个矩阵与表 3.2 形式完全相同，

表 3.3

	合作 C	背叛 D
C	R	S
D	T	P

但是要成为囚犯困境还必须满足两个条件：

$$T > R > P > S$$

$$2R > T + S$$

两个战略是合作 C 和背叛 D。囚犯困境推翻了一种传统的理论。这种理论认为，一群有共同利益的人一定会采取追求这个利益的集体行动，即如果每个人最大限度的追求个人利益，则集体利益也就是个人利益的总和会自然达到最大。美国著名政治学家奥尔森 (Mancur Olson) 在 1965 年发表了经典性著作《集体行动的逻辑》，探讨各利益集团的集体行动的微观基础。在 1971 年著名美国政治学家哈丁 (Russell Hardin) 进一步指出，奥尔森所描述的集体行动问题在结构上等价于“n 人囚犯困境” (*n-person prisoner's dilemma*)。由此可以推出一个结论，为了维护全集团中所有个人的最大利益，必须存在一个有强制性的绝对权威，以对抗个体进入 ESS 即背叛战略的自发倾向，把博弈维持在合作战略上。这种权威可以是一个国家机器，也可以是一个被全体接受的

基于形而上学宗教信仰的道德体系。用这种理论可以解释很多现存的社会形态和道德的价值，使之成为可计算的量，但是由此又会引出一些用这种理论所难以解释的问题。第一个问题是权威从何而来的问题。按照这种理论，似乎应该是在一个人人背叛的 *ESS* 状态中，突然出现了一个纵观全局的有远见卓识的圣人或上帝的代言人，说服或压服了所有人接受其观点，建立了社会的秩序。但是从认识论的角度看，在人人背叛的环境中，圣人也难以认识合作的好处，而且即使出现了圣人，其他人也没有理由接受其观点。可以想象在大家乱挤的洪流中有一个人站出来大喊“排队！”是不会收效的。而且不论是孔子还是耶稣基督或释加牟尼，当他们在世时也都不处于具有强制能力的权威者地位。第二个问题更为明显，在动物群体中不可能存在道德约束或远见卓识，在肯定没有权威者来灌输或坚持某种道德的条件下，自私的基因为什么会导导致互惠利他的合作行为的进化？

阿克赛洛德 (Robert Axelrod) 在 1979 年所进行的研究解决了上述的难题，并写成了《合作的演化》一书，引起了轰动。这一研究成果给出了道德可以自发产生的具体条件，也说明了动物互惠利他行为的理由。阿克赛洛德的研究是从实验开始的，他设计了一个计算机对战的战场，请博弈论专家们各自设计了一些战略，让各个战略分别和其他战略（也包括自己）进行囚犯困境的多次对战，最后看哪种战略累计得分最多。所有战略都由合作和背叛两种战略组成，可以根据过去与某个对手对战的经历来选择这次应该与之合作还是应该背叛。此次实验一共收集了 14 个战略，其中最简单的战略程序只有 4 行，最大的 77 行。每场比赛每个战略与其他战略各赛 200 次，共进行了 5 场。参战的是这 14 个战略再加上“随机战略”，即随机决定合作和背叛的战略。在比赛中获得优胜的是研究囚犯困境的名人，多伦多大学的心理学家兼哲学家阿纳托·拉帕波特 (Anatol Rapoport) 的战略。是各战略

中程序最短的一个，名字叫礼尚往来战略 (*TIT FOR TAT*, *TFT*)，这个战略惊人地简单：第一次相遇时合作，从第二次以后采用上次对方采用的战略。这个战略的行动原则很简单，无非是一报还一报，即“以其人之道还治其人之身”。阿克赛洛德分析了 *TFT* 取胜的原因，总结出两条取胜所必须的重要性格：第一是要当“君子”，决不主动背叛对方。第二是要“宽容”，对于对方的背叛只报复一次，如果对方重新合作，则我方也转为合作，决不记仇，冤仇宜解不宜结。*TFT* 正是具备了这些性格，所以能够取胜。阿克赛洛德发现这些性格的重要性远远超过想象，通常人们都认为“义”是“利”的制约条件，这个结果却表明“义”正是取“利”的法宝。他考虑在这个分析的基础上还可能创造出更好的战略，于是组织了第二次会战，在计算机杂志上出了征集广告，并且给每个参加者寄去了上次会战的分析结果，指出哪些作法可取，哪些作法不佳。这次会战有 6 个国家 8 个学科的人参加，其中有 10 岁的少年，也有著名学者，最短的战略程序仍是 *TFT*，最长的有 125 行。这次参加会战的有 63 个战略，比上次作了精心的改良。出乎意料的结果是，得分最高的仍是 *TFT*，而且处于前 15 名的战略基本上都是“君子”，只有一个“小人”得了第 8 名。而在倒数 15 名以下的只有一个“君子”。在这以后又设计了一种进化论的实验，即在会战开始时持各种战略的“人数”相同，随着对战的进行，得分高的意味着适应度高，所以其“子孙”会增加，得分少的会被逐步淘汰。在经过 1000 次的对战之后，*TFT* 仍是第一，而且与第二名的差距越来越大。有趣的是，如果是一种战略对一种战略，*TFT* 并不是常胜将军，对别的战略经常失败，顶多打个平手，例如 *TFT* 对“全面背叛”战略，第一次显然输，从第二次开始打平手，累计起来还是输。*TFT* 的优越性是能够对付多种战略，而且善于在合作中与对方分享利益（而不是总想独吞！）。进一步总结取胜的经验，除了原来的“君子”，即“害人之心不可

有”，决不主动背叛，以及“宽容”，即适度报复，“防人之心不可无”，但也不记仇之外，又发现了两条原则，其一是“性急”，即遭到背叛之后尽快报复，有一种战略与 *TFT* 大体一样，只是对方背叛两次才开始报复，这个战略在第一次会战中成绩不坏，第二次的会战中名次落到了后面。另一条原则是“透明”，即容易被对方理解，这个原则好象与常识相反，一般总认为城府很深的人会上算，其实不然，有些战略设计得十分复杂，使对方无法理解，只能当成“不知好歹”的战略处理，而对付“不知好歹”的战略，你合作他却未必回报，最佳战略是“全面背叛”，结果不透明是会吃亏的，这也就说明了“大智若愚”的原理。象“随机战略”是最不容易被对方摸清楚的，表现出典型的“不知好歹”，结果战绩极差。“透明”原则表明，即使在一个“尔虞我诈”的环境之中，从长远看，“诚实”也可以带来利益。在自由竞争之中最终能发展的还是讲信誉的企业。后几次会战结果仍然表明，“君子”即不首先背叛的原则比我们想象的要重要，有不少“耍小聪明”的战略都由于违反这一原则而遭到惨败。例如有的战略程序在 *TFT* 的基础上加了一条，以 10% 的概率主动背叛一下，想趁人不备时偶然捞一把，还有的战略程序为了试探对方的报复能力而主动背叛一次，结果这些破坏相互信任的举动结果都是得不偿失，付出了高于利益的代价。

在一系列会战实验之后，阿克赛洛德对于 *TFT* 是进化稳定战略给出了严格的理论证明，并且求出了其条件为：

$$W \geq \frac{T - R}{R - S} \quad \text{且} \quad W \geq \frac{T - R}{T - P}$$

式中的 W 是下次还要与对手进行对战的概率。如果只进行一次对战， $W=0$ ，如果对战总次数为 N ，则 $N=1/(1-W)$ 。当 N 充分大时，可以认为 $W=1$ 。其他字母的含义见表 3.3。显然，只要满足表 3.3 成为囚犯困境的两个条件，且 $W=1$ ，即对战充分长，则

上述条件也必然满足，即 *TFT* 是 *ESS*。如果只进行一次对战，当然背叛是最佳战略。

阿克赛洛德还对 *TFT* 战略在不同阶段的性能进行了分析，研究 *TFT* 为什么能够从无到有，从小到大。第一个问题是，个别的 *TFT* 能否侵入自私的海洋。由基因突变产生的 *TFT* 在一开始肯定是极个别的，在其他个体全都是采用全面背叛战略的不合作个体时，*TFT* 显然不利，面临被淘汰的危险，首次合作失败导致适应度下降，如果这个 *TFT* 个体在自私的大海里大范围活动，多次碰到新对手而被坑害，当然会被淘汰，但是如果他只在小范围活动，只和少数自私者打交道，那么他很快会适应环境，在多次对战中平均损失不大，具有和自私者近似的适应度。只要 *TFT* 有了子孙，构成一个小集团，就可以通过合作而受益，提高适应度，使小集团扩大。也就是说，*TFT* 是可以自发产生的。第二个问题是，*TFT* 能否应付各种复杂局面，对付各种不同战略。这一点已经得到证明，由 *TFT* 具备“君子”、“性急”、“宽容”、“透明”等优良品质，使其既能有效地对付全面背叛的自私者，又能不失时机地从合作中得到利益。第三个问题是，全部由 *TFT* 组成的集团，能否抗拒其他战略的侵入，由于和第二个问题相同的原因，这一点也是肯定的。而且从集团之间竞争的角度看，*TFT* 集团显然胜过全自私集团。总的来说，只要合作是有利的，象 *TFT* 这样的礼尚往来的“君子”就可以在“小人”的海洋中自发地产生、壮大，并最终淘汰掉“小人”。而全由“君子”组成的集团，即使在“小人”成群结伙地大举侵入下，也能岿然不动，具有极高的稳定性。用阿克赛洛德的说法就是：“社会进化的齿轮已经装有防止倒转的棘轮”。在这个过程中，外加的权威是完全不需要的。

阿克赛洛德所得到的结论似乎与哈丁的结论相反。应该注意的是，“君子”自发地产生比起哈丁所研究的情况要求更多的条件。其关键在于对于背叛行为有可能进行报复，这就要求博弈要能在

相同对手之间重复多次。这种情况在社会中相当于熟人之间的交往，在不具备个体识别能力的生物中则相当于固定的地点、对象或有血缘关系的集团内部。因此在公共汽车站经常看到的“乱挤”现象，在一个单位集体乘车旅游时不会发生。这就是熟人与生人的区别在社会中普遍存在的理由。在一个相互都认识的小村落中，淳朴的民风完全可以不用权威来维持，众人的唾沫可以淹死反道德分子，但是在一个谁也不认识谁的大城市中，领导者或道德的权威就是不可缺少的了。但是追本溯源，道德和领导者还是在自发地产生之后，逐步获得其外在强制性权威地位的，圣人并不是凭空创造了道德规范，只是总结了已有的行为准则而已。

无论是哈丁的分析还是阿克赛洛德的分析，其基本问题都是囚犯困境，即共同的前题是合作的总体利益大于相互背叛或单干。以我们熟悉的两种动物为例，狗是比较“忠诚”的，而猫是“奸臣”。这是由于狗与猫的食性不同，犬科动物通常捕食比自己更大的动物，象狼群要通过集体围猎的相互合作才能捕到鹿，然后分而食之；而猫科动物则是单独捕食比自己弱小的动物，如果旁边有别的猫，只是增加一个争夺的对手，毫无利益可言。因此狗的合作的性格在进化中得到发展，表现出某种为其整体协调而克服眼前利益的能力，即采用合作而非背叛的态度。我们很容易训练狗不吃别人给的东西，或面对食物等待人发出许可才吃，但要想训练猫不去偷吃我们没有收好的食物完全是徒劳的。这就是“猫性”与“狗性”的区别。人作为一种社会的动物，在分工与合作之中获得巨大利益，具有极强的“狗性”，所以在社会中能够自发地产生道德的行为准则，并且随着社会规模的扩大，成为多数人相互不认识的集团时创造出外在的强制性道德权威。这种道德权威往往会带有某种形而上学性，具有天然的对抗批判的性质，如果不具备这种性质，每个人都可以用自身眼前利益来否定道德，那么道德就会失去作用，不成其为道德了。但也正因为这种对抗批

判和分析的性质，使得道德具有危险性，在道德体系之中完全可能藏有并不符合全局长远利益的成分而难以被人发现，从而导致这样的可能性：“在通往地狱的道路上，铺满了善良的愿望。”

3.4 广义进化论的软硬结构模型

地形图模型可以表现固定条件下参数的寻优过程，博弈论模型则进了一步，可以表现在固定条件下方法、战略、行为模式乃至价值观的择优过程。但是对于寻优和择优的结果怎样影响其后的条件变化，则缺乏一个有效的更深层的模型。如果把整个的进化过程看成一个链条的话，那么地形图模型和博弈论模型所表现的就是链条的某个环节，而对于各环节之间的接合部，对于全链条成长过程则缺乏整体认识。就如同我们看一个微分方程可以直接了解局部的和短时间的变化倾向，但是如果不求出微分方程的解，就很难对全局的长时间的整体形象有所了解。从这个角度看，达尔文的进化论可以说是进化论的微分形式，目前关于神经网络学习的理论也可以看成是学习理论的微分形式。要突破这种理论限制，不是靠简单的积分过程可以作到的，必须在理论中加入自我表述的因素。本节通过建立软硬结构模型，把自我表述的机制引进达尔文进化论，从而推广出广义进化论。

达尔文进化论的核心机制是变异与选择，广义进化论在此基础上增加的机制是软硬结构的相互作用。对于自我表述系统，可以分为两个部分，如图 3.3 所示。一部分叫作软结构，是可变的；另一部分叫硬结构，是不变的。软结构与硬结构之间相互作用，硬结构对软结构的作用是支持，软结构对硬结构的作用是建构。硬结构执行系统的“日常”功能，保证系统的生存（存在），软结构在硬结构支持下对硬结构进行建构，使整个系统不断进化（演化）。在图中软结构用椭圆表示，硬结构用矩形表示。

以上是对软硬结构模型的高度抽象化的描述，为了理解“软结构”、“硬结构”、“支持”、“建构”这些基本概念，以下我们用一些具体系统作为例子进行形象化的说明。

第一个例子是树木的成长过程。成长不是进化，但这个简单的例子有助于我们理解软硬结构模型的基本概念。在树木成长过程中，芽的顶部存在着一个生长点如图

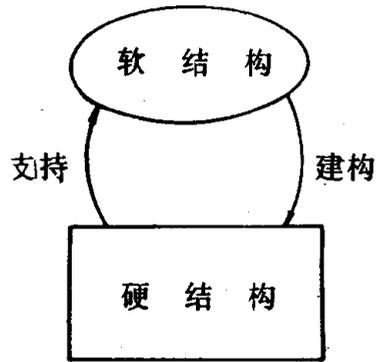


图 3.3

3.4a，生长点处的细胞能够分裂增殖，成为树木成长的软结构。分裂出来的细胞逐渐长大、伸长、依其位置分化成树枝、树皮、树

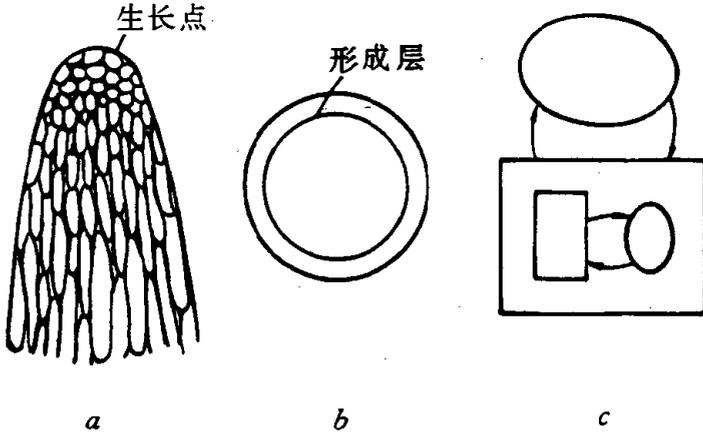


图 3.4

叶、树根等多种组织，硬化成固定形状，这就是软结构对硬结构的建构。各种定形的组织是树的硬结构，完成吸收、制造和输送养料的日常功能，而这些养料提供给生长点，满足其进行细胞分裂所需的物质和能量要求，这就是硬结构对软结构的支持，这种支持还表现在树干和树枝对生长点进行机械支撑，使生长点固定于

一定的几何位置，而通过生长点的建构，树的硬结构逐渐长高，生长点也随之升高，这就描绘了软硬结构及其相互作用所形成的环状结构通过自我表述实现“自举”的过程。如果把人工智能比喻为“爬树”，那么人的智能就是这棵不断成长的“树”，认知科学就是要弄清这棵树是怎样自己长高的。

如果分析得更仔细，可以发现在树干或树枝这种硬结构之中存在着下一个层次的软硬子结构。图 3.4b 是树干的断面图，在树干中有一圈形成层，是软结构，其中的细胞也能分裂，向内部分裂的细胞硬化成木质部，向外的增生长成韧皮部，也是软结构对硬结构的建构。这个子结构使树木在长高的同时能够长粗，这样才能保证树在长高时不被自重压倒，有一定的抗风能力。总起来看软硬结构与子结构之间形成一种“递归”或“嵌套”的关系，可以构成多重的复杂结构，如图 3.4c 所示。象竹子这样的植物，由于没有形成层，不能越长越粗，所以达到一定高度就只能停止生长，以免折断。

第二个例子是用滑模工艺来说明建造人工建筑物时的自举过程。通常在建造混凝土建筑物时，需要先搭脚手架、立起重机，然后支模板、浇灌混凝土。建筑物很高时，搭脚手架的成本增加，起重机的稳定性也变差。这时可以采取半自举的措施，由已盖好的楼层支持脚手架，随着楼层升高而升高起重机的位置，并把起重机的腰部靠已完成的楼层进行加固。对于形状简单的建筑物，例如圆筒状的高塔，可以采取如图 3.5 所示的滑模工艺，可以完全省去脚手架和自立的起重机，实现全部的自举过程。图 3.5 是滑模工艺的示意图。浇灌混凝土用的滑模由两个环状的模板，即外模板 A 和内模板 B 组成，内模板 B 上固定着工作平面 C，平面上装有小型吊车 D，吊车把装有混凝土或钢筋的小车 E 吊上平台，在平台上焊接钢筋并把小车中的混凝土浇灌到模子里去。内外模板 A 和 B 下方经由油压千斤顶 G 连接着卡具 F。当滑模被混凝土灌满

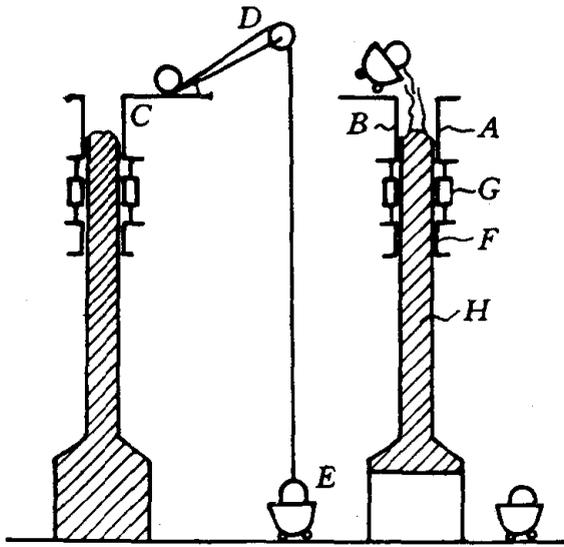


图 3.5

时，滑模下半部的混凝土已经凝固，此时收紧卡具 F ，使其固定在已固化的混凝土壁 H 上，然后开动油压千斤顶，把滑模顶起，使其向上滑动一段距离，上部再次出现空模，以备浇灌，其后松开卡具 F ，并操纵油压千斤顶收缩，提升卡具 F ，准备下一次顶升滑模。

在滑模工艺之中，滑模及其附属设备是系统的软结构，而建造中的混凝土建筑则是系统的硬结构。混凝土在滑模中的固化过程是软结构对硬结构的建构，而滑模依靠混凝土壁的支撑进行爬升则是硬结构对软结构的支持。所建的塔的高度只受最下部混凝土抗压强度的限制。滑模工艺的例子非常单纯，自举过程所得到的只是几何高度，还体现不出任何创造性，但是可以使我们得到自我表述系统的最基本的印象，即提着自己头发飞上天具体要用什么样的方法。

第三个例子计算机的进化过程。首先要提起注意的是，计算机中的软件与硬件之分，并不对应着软硬结构，因为软件与硬件

之间存在着逻辑等价关系，在目前来说，计算机中的软件和硬件都属于硬结构，软结构是指计算机软硬件的开发者，也就是说，我们把计算机及其开发者看成一个系统来讨论问题。当然这并不意味着在计算机的软硬件上不能实现软结构，将来我们弄清了实现软结构的方法，就可以在计算机内造出软硬结构，用这样的计算机就可以自己给自己开发软件。本书的目的之一，就是弄清在机器上实现软结构的各种必要条件和一些方法。但是从目前的现状出发，还是先讨论人机系统比较合乎时宜。

把计算机的软硬件看成硬结构，把开发者看成软结构，那么开发者所开发的软件工具添加到计算机的软件系统之中，就是软结构对硬结构的建构；而日益丰富和高级化的开发工具和软件环境，又对开发者的开发工作给以越来越强的支援，逐步提高开发效率，实现硬结构对软结构的支持。计算机系统特别是软件系统的进化史，就是这样一个自我表述的自举过程。初期的计算机只能靠机器语言编程序，此时软件开发是很困难的，从总体设计到算法设计，再到程序设计，然后还要由人工把程序写成机器语言，安排存储地址，此后才能把程序输进计算机，进行实际运算。计算机的奇妙之处在于，无论什么工作，只要能够形式化成为死板的手续，都可以由计算机自动完成。因此很快人们就把一些任何程序都需要的公共部分，例如输入输出，外部设备控制和管理，机器内部资源管理等编成统一共用的监控程序或操作系统(OS)；以及把程序翻译成机器语言这样的重复繁琐而又有一定之规的工作改由机器自动完成，从而创造了汇编语言和汇编软件。这样就大大提高了计算机对于开发者开发软件的支援能力，使其后软件开发能以更高的效率进行。沿着这条自举的路线，出现了各种高级语言编译软件，程序调试工具软件，高效易懂的图象人机界面，计算机辅助软件工程(CASE)等等。如果离开用软件工具编软件工具的自我表述原理，离开进化产生的高级软件工具，仍用机器语

言来编程的话，现今的大规模软件生产是难以想象的。即使在微观的个别软件编制过程中，有时也能利用自我表述的进化过程。例如我们要从无到有地编制一个高级语言（例如 C 语言）的编译软件，而手头只有汇编软件作为工具。比较笨的办法是用汇编语言从头到尾地写出编译软件的源程序，然后用汇编软件把这个源程序翻译成可执行的机器语言程序。这个方法效率极低且移植性差。最好是用 C 语言去写 C 的编译程序。但是在我们还 没有 C 语言的可执行编译程序时，用什么方法能把 C 语言写成的 C 编译程序的源程序翻译成可执行的 C 编译程序呢？这里就要用到自我表述的进化方法。首先我们用汇编语言写成一个 C 编译程序的子系统，这个子系统中只要求必不可少的极为贫乏的编译功能，因此可以作得很小。然后用汇编程序把子系统的源程序翻译成可执行程序，就得到了一个“小 C”的可执行编译软件。有了“小 C”作为工具，就可以对小 C 逐步进行功能扩展。每次都 用前一个版本得到的工具进行开发，得到功能更强的新版本的工具，最终得到功能完全的编译软件。在今后的版本升级时也是用相同的方法。

计算机系统的开发是一个创造过程。创造主要是由开发者完成的，但是如果没有人 与机器组成的软硬结构，没有通过开发增加机器功能和机器功能支持开发这种软硬结构的相互作用，开发者就不可能进行大规模的创造。软硬结构的关键作用在于把无数小创造积累成大创造，其基本的机制是自我表述，即创造的结果能支援创造过程，而创造过程又产生新的创造结果。对于这种把小创造积累成大创造的机制的重要性我们往往是认识不足的，就象对于计算机系统的开发会有这样的观点：所有的创造性工作都是由人（开发者）完成的，人机关系并不包含创造的因素。产生这种看法是由于对人的智能缺乏分析态度，人所表现出的看起来与机器有质的差别的创造性，起源于人脑内部的软硬结构，即把更小的创造积累成大创造的能力。关于这一点将在下一章详细讨

论，在这里我们可以暂时用一个递归的模型加以解释，即在人与机器的软硬结构之中，人这个软结构内部包含着次级的软硬结构，这样逐步把大创造分解下去可以得到任意小的创造，而实现小创造是非常容易的，掷一次骰子也可以说是一个小创造。进化的最基本动力就是掷骰子，即随机变异，而经过几十亿年的积累能创造出象人这样具有创造性的生物。由此可以体会到创造的积累过程何等重要。要想直接产生大创造，就会碰上指数爆炸的困难，而指数爆炸是对规模而言的，指数爆炸的问题在规模足够小时并不构成困难，这就是小创造容易实现，而大创造必须依靠积累的原因。光靠阿希贝的原理是走不了多远的。

第四个例子是社会。社会是由众多的个人组成的具有复杂结构的自我表述系统。在社会这个大系统中，我们可以发现很多软硬程度不同的子系统及软硬结构，而每个人又不同程度地参与到软结构和硬结构之中。孟子曰：“仁者乐山，智者乐水”，把人分为“仁者”和“智者”两类。仁者是指参与硬结构的人，这些人的性格是“乐山”，即喜欢不变的东西，“山”是稳定性和不变性的象征。反之，“乐水”的“智者”则是组成社会软结构的人，其性格倾向于喜欢流动变化，象水那样。官僚可以说是“仁者”的典型，而思想家，发明家是典型的“智者”。当然，一个完整的人通常总是具备乐山与乐水两种性格，只是由于社会分工不同，在对社会的作用方面表现出不同偏重而已，一个人可以在本职工作中当个仁者，而在业余生活中是个智者，如同鹰鸽博弈中的混合战略一样，既可以是按一定的百分比由纯鹰派和纯鸽派组成，也可以是每个个体按一定的概率采取鹰鸽两种战略。乐山的仁者容易被看成是保守派，乐水的智者通常被看成是革新派。人们往往把这两种社会组成成分看成是对立的两极，而看不到这两种成分组成社会的软硬结构，以及相互之间支持与建构的关系。正由于这两种成分的作用，社会才得以存在和发展。官僚与思想家具有

完全相反的性格特征。“官僚”这个词给人的印象是：循规蹈矩、照章办事、因循守旧、对上唯唯诺诺、对下独断专横、尊重法律、权威、传统、制度、道德，宁可犯路线错误、不能犯组织错误。“思想家”则是反其道而行之，不承认任何神圣不可侵犯的存在，一切现存的结构都在其思辩、批判的解剖刀之下，以提出人所未知的惊世骇俗的看法为己任。这两种性格是其在社会中所处的不同结构性质使然。社会的硬结构要求其成员也具有充分的僵硬性，就如同马克思把国家机构称为“国家机器”，其中的零件如果变软，机器是不能正常运转的。路线错了可以由机构自身的功能修正，组织如果瘫痪或崩溃了，什么路线也执行不了。官僚如果具有思想家的性格，迟早丢了乌纱帽。而思想家如果按官僚的思维方式考虑问题，则什么新思想也不可能产生了。思想家作为社会的软结构，要通过建构作用对社会产生影响，这种建构是一个比较复杂的过程。思想家的任务是产生新思想，相当于生物进化过程中的“变异”。变异不一定是有利的，据生物学家的说法，变异绝大多数都是有害的，有益的变异只占其中的 10^{-8} 即一亿分之一。思想家所创造的思想不是随机的胡说八道，而是经过深思熟虑即大脑内部的选择过程，因而好思想的比例可能大得多，但也不可能都是好的，还要经过进一步的选择，但只要思想是新的，就能算得上一个变异。为了经受选择，思想家要把新思想写成文章或著作发表，或者通过讨论或演讲的方式口头发表。如果这种思想不引起人们的兴趣和注意，或者虽然引起争议却未能被人接受，那么这种思想就在选择中被淘汰。如果新思想通过了这个选择，而被多数人接受，经过多数人的实践觉得有效，但是与现行法律法规有发生冲突之处，那么这种思想就要反映到与立法过程有关机构之中，通过立法程序，修改有关的法律法规，就完成了软结构对硬结构的建构。此时官僚们仍是照章办事，只是“章”本身发生了变化，在另一方面，软结构的存在和软结构对硬结构的建构

都离不开硬结构的支持。这种支持可以说是无处不在：思想家作为一个活人，其生存必须依靠社会的行政、商业、工农业、交通、通讯等各种现存硬结构的支持。新思想不可能凭空创造，须要依靠图书馆、文献检索、大众传媒等各种硬结构提供信息源。新思想的传播受到法律所规定的言论、出版自由的支持。新思想对法律这种硬结构的建构也要由法律所规定的立法程序来完成。古代比较原始的法律大多对立法程序不作规定，而现代的法律、章程，乃至合同，都有关于修改其自身的有关条文，也就是说都有自我表述的功能，以支持软结构对其自身的建构。思想家在思想上可以批判任何权威，否则不可能产生全新的见解，但在行动上必须在法律范围内活动，以换取硬结构的支持。当然，这里所说的“官僚”和“思想家”只是一种典型化的概念，处于社会中的每个人都在不同程度地执行软结构和硬结构的功能，每个人的头脑中都会产生“变异”，每个人在行动时都不断面临着“选择”，每个人都受到社会环境的制约和支持，每个人也都对社会造成影响和建构。社会就是靠着这种复杂的相互作用存在和演化的。

第五个例子是科学技术的发展。科学是各种文化现象之中最具增殖性的一种，其发展和创新的速度高于其他任何一种文化现象，几百年现代科学的发展对于人类文明的贡献，远远超过几千年文明史所积累的文化财富的总和。现代科学技术高速发展的一个重要原因就是其体系内部形成了相当完善的软硬结构。在科学范围之内，科学的已知领域是硬结构，已知领域和未知领域的交界处是类似于形成层的软结构，随着科学的发展，已知领域逐步扩大，形成层向前推进，发现更多的未知领域（问题）。科学的边缘地带，是最活跃的创造区域。科学的软结构具有高于一切文化现象的柔软性，这表现在科学的一条基本原则，即“干什么都行”和“怎样干都行”，不存在科学不可问津的禁区，不存在科学所不能采用的方法。科学欢迎“标新立异”，不怕“异端邪说”。科

学论文的价值很大程度上决定于这篇论文是否提出了与众不同的理论或方法，对众所周知的理论进行简单的组合在学术上是没有太大价值的。科学软结构的这种“求异”倾向的价值观使得科学领域中得以诱发大量的“变异”，为创造提供了丰富的可能性。科学软结构的另一个优越之处是在科学领域中存在着稳定的、明确的、公认的价值观，使我们能够很容易地对各种“变异”进行选择。科学之所以不怕“异端邪说”，就是因为有了这样一组价值观作为过滤器，我们可以保证坏的理论不会混入硬结构之中去。总是可以在现有变异之中选出最好的予以保留，把坏的淘汰掉。通过这样的选择，改进，充实，把新的成熟理论纳入到科学的理论体系之中去，就实现了软结构对硬结构的建构。科学的硬结构，即成熟的科学体系具有相当高的硬度，要想推翻一个成熟的科学理论是非常困难的。这种硬度不是靠某种外部权威来维护的；与此相反，科学理论的硬度恰恰是靠欢迎推翻来实现的。世界上不存在任何权威机构来审定一种科学理论是否具有不可推翻的真理性，或发布命令对某种理论实行保护。反之如果有人能够推翻一个公认成熟的科学理论，将会在科学界获得巨大的荣誉。这是一种破坏性实验的原理，如果你想证明某种玻璃杯是摔不破的，最好的方法是设置一项奖金，谁能摔破这个杯子就给一笔大奖。在理论上任何密码都有可能被破译，曾经有一个通讯部门设计了一种密码，不久就有专家指出这种密码在理论上是不安全的，于是通讯部门发出巨额悬赏，在数年中无人实际破译成功，用户对于实用的安全性也就放心了。以上所说的是理论的相对硬度，在理论的预见力方面还存在着更带绝对性的硬度。例如牛顿力学。在相对论取得公认成功时，牛顿的绝对时空观受到致命的打击，但是牛顿力学在其适用范围之内的预见力并未被破坏。可以认为，不论未来物理学怎样发展，牛顿力学的预见力被破坏都是不可想象的。假如我们想象万有引力突然消失，地球上的人类都会失去其

生存的可能性，那么这种想法本身就无处存身了，变成无意义的东西。在软结构建构之下不断增长的硬结构又对软结构起到支持的作用，这种支持表现在“变异”与“选择”两个方面。“变异”指的是新理论的产生。理论是不能从实践中自发产生出来的，实践对理论只能起到验证和提出问题的作用。新的理论来源于对旧理论的组合和加工，硬结构越大，旧理论及对理论的加工方法越丰富，产生新理论的条件也越充分。由于科学理论体系的内部一致性，有些在一个学科中得到的有力的方法和模型可以泛化到多学科之中，产生普遍的意义，这就是库恩所说的“范式”的作用。从“选择”的角度看，硬结构所包含的理论越丰富，在选择时对新理论所施加的一致性约束也就越严密，选择也就越有充分的依据。科学的硬结构对软结构的支持还表现在随着硬结构的扩大，软结构逐渐向前推移，科学的研究重点不断向着更深刻，更广泛，更有预见力的方向移动。

科学体系不仅在其内部具有软硬结构，而且在与其他文化现象之间也存在着软硬结构的关系，例如在科学与技术、科学与教育的关系中都可以发现软硬结构。在科学与技术的关系之中，科学是软结构，技术是硬结构。科学研究中取得的新成果引发技术上的革命性进展，实现建构过程，而技术的发展又为科学研究提供越来越先进的实验手段，对软结构给以物质支持。就科学与教育的关系而言，科学是软结构，教育是硬结构，科学对教育的建构依靠的是科学的明晰性，即在科学研究之中历尽千辛万苦而得到成果却可以在几个小时的授课之中让学生掌握，因此可以把几千年科学发展的结晶压缩到几十年的教育内容之中，让下一代人得以迅速进入科学的前沿，让科学事业得到源源不绝的人力资源的支持，保证了新一代的科学家总能站在巨人的肩膀上，从新的起点起步。这种硬化建构的过程与几十亿年的生物进化过程压缩到一个生物个体发育过程之中的机制是完全相同的。这样的一系

列的软硬结构的相互作用使得科学体系发展成为各种文化现象中最为庞大的理论体系，而且是增长最快的理论体系。科学地球不仅庞大，而且膨胀速度也是最高的。

以上的一些例子可能会引起一种错觉，似乎软硬结构是无处不在的。下面以两个不包含软硬结构的文化现象作为例子，从反面加深对软硬结构的理解。只有硬结构而没有软结构的典型例子是宗教。宗教是一种封闭的理论体系，具有顽强的自我保护性质，其基本教义是不容变更的，因此从价值观上来说是不欢迎变异的，排斥和打击一切与教义不符的异端邪说。从选择的机制来看，虽然每一种宗教都具有确定的价值观体系，但是总的来看是没有通用性的，因此找不到一种公认的价值观体系来判断哪一种宗教或哪一种教派的理论更好。由于宗教本身的抗拒变化的倾向，对其进行建构几乎总是违反教义的，在宗教中不存在修改教义的合乎教义的手段。因此宗教的硬结构不支持软结构，也不支持软结构对硬结构的建构。与宗教相反，艺术是一种只有软结构而没有硬结构的文化现象。艺术是经常产生变异的，但是在艺术领域中不存在稳定的、明确的、公认的价值观，对于不同的艺术形式和艺术风格，通常是难以断定其间优劣的。因此艺术是难以形成硬结构的。艺术不具有象科学那样的明晰性，一代艺术大师的成就并不容易成为下一代艺术家的当然起点，艺术巨人的肩膀是不那么容易爬上去的。因此在艺术领域中很难形成可以依靠建构而不断增长的硬结构。

以上的例子并不是说宗教与艺术是不进化的。作为文化现象，宗教、艺术与早期未形成大体系的科学一样，都处于由变异与选择组成的文化进化的洪流之中。宗教由于教义的变异会产生不同的教派或分裂成不同的宗教，各自拥有信徒的数量是人们选择的结果。在艺术领域中也就会不断地变异出新的艺术风格，而某种艺术风格能否流传则取决于人们基于某种偏爱的选择。早期的科学

技术也是由一些不成体系的零散发明和发现所组成，也是缺乏可被建构和提供支持的硬结构的，类似于小行星的状态。中国古代的科学技术可以说是这种非体系科学的典型例子，由于缺乏建构机制科学技术只是零散发明和发现的堆积，自生自灭，有些技术不得被反复发明多次，有些技术则完全失传了。拥有发达的软硬结构的现代科学体系与这些普通的文化现象相比，不只表现出很高的进化速度（产生新知识的速度），而且具有普通的文化进化所没有的“进步保证”。这种进步保证依赖于科学的明晰性，使我们总能站在不断长高的科学巨人的肩膀上；也依赖于稳定的价值观体系，使我们能有把握分出科学理论的高下。这种进步保证就表现在：任何当代的物理学家都可以说他对牛顿力学的理解比牛顿更深刻，因为牛顿不知道相对论和量子力学而他知道。任何当代的数学家都可以说他对欧氏几何的理解比欧几里德更深，因为他知道非欧几何而欧几里德不知道。任何一位当代的进化生物学家都可以说他对进化论的理解比达尔文更深，因为达尔文既不知道孟德尔—摩尔根的遗传学，也不知道分子生物学，因此他虽然为进化提供了大量证据，但在理论上的深度远不及现代进化论者。因此在科学的领域中永远不会出现复古思潮，后人比前人高明是理所当然的。不具备软硬结构的文化现象则不具有进步保证。无论是宗教还是艺术领域中都会出现复古思潮。齐白石的弟子不敢说自己的画技超过齐白石，梅兰芳的传人不能说自己的艺术水平盖过梅兰芳，现代雕塑家没有理由认为自己的艺术造诣肯定高于古希腊，当代小说家也不能保证自己的作品强过红楼梦。现今的高僧不能说自己比释伽牟尼更懂得佛教，当代的教皇也不敢说自己比耶稣更理解基督教，现代的儒学家也没有理由说自己超过了孔子，虽然在孔子之后又有了程朱理学，但是程朱理学对于儒家思想来说到底是发展还是倒退和背叛是一个大可商榷的问题。

对于以上一些例子的分析可以说明，分析一个自我表述系统

可以用软硬结构模型作为基本思路，而构筑一个具有软硬结构的能够有效地执行其功能而具备进步保证的自我表述系统需要满足一系列的条件。其中最基本的是四条：

第一，系统包含硬结构。

第二，系统包含软结构。

第三，硬结构支持软结构。

第四，软结构对硬结构进行建构。

为了实现有效的软硬结构及相互作用的功能，对于以上四点还有进一步的要求。这些要求可以概括为四句话：软硬分离，硬硬软软；增长支持，高层建构。软硬分离，硬硬软软的意思是，由于软结构与硬结构在系统中所起作用不同，其要求也是相反的，所以只有对两种结构进行充分分离，避免混淆，才能使两种结构各自处于最佳工作状态而互不干扰、互不妨碍。硬硬软软这个句子的语法结构与“君君臣臣、父父子子”一样，“君君臣臣”是指皇上要象个皇上的样子，大臣要象个大臣的样子，各司其职，这是适用于任何一种组织结构的普遍要求。“硬硬软软”的意思是硬结构要象个硬结构的样子，要充分的硬，而软结构则要充分的软。硬结构在系统中所起的作用是维护系统的存在，对其要求是高稳定、高效率（高经济性）、高预见力。象法律规章、组织机构这类的硬结构，必须尽可能的稳定，法律或制度如果朝令夕改，则让人无所适从，助长短期行为。一个办事机构如果无法可依，无章可循或者有法不依，法制硬度不够，遇事都要“研究研究”再定，则肯定是没有效率可言的，也是极不经济的。要提高法律制度或组织机构的稳定性，必须使其具备一定深度的理论支持，减少“就事论事”的成分，从而具有较好的预见力，当出现新情况时，不用靠发布暂行条例或建立临时机构来适应。软结构与硬结构所要求的性质恰恰相反，为了实现软结构在系统演化中的作用，软结构必须充分的软，即尽可能实现大量的变异，处于极不稳定的状

态；大量的变异会是不好的变异，在选择中被淘汰，因而浪费很大，经济性极差；演化作为一种富含创造性的过程，从定义上就是不可预测的，可预见的东西不能称之为创造。从以上几点看，似乎软结构的评价准则完全与我们的善理三大原则相逆反，但是从大的时间尺度上看，正是软结构的存在，使得硬结构有可能在符合评价原则的建构之下，逐步发展得更稳定，更经济、更有预见力。即使是同一个体在组织中起到软结构和硬结构两种作用时，软硬结构的分离也是重要的。例如在“集思广议”的过程中，“集众思”是组织软结构的功能，组织成员在这个阶段应按软结构的要求，尽可能充分发表不同意见，实现大量变异；在经过选择形成决议之后进入“广忠议”的阶段，组织中的成员应按硬结构的要求，在行动上不折不扣地执行决议。只靠软硬结构的分离还不足以实现进步保证，在软硬结构的相互作用方面，还必须满足“增长支持，高层建构”的要求。增减和高低表示“进步”的方向性。在讨论进步的概念时，向什么方向的演化算作进步必须是明确的。有了明确而稳定的价值观，支持是否增长，建构是否在高层，就成为可判断的了。在树木长高和滑模工艺的例子中，我们把几何的高度，距离地面尺寸的增长作为一种明确的价值，那么随着硬结构在建构过程中增高，作为软结构的生长点和滑模也相应增高，就满足了增长支持的条件，而建构过程总是发生在树和塔的高端，也就满足了高层建构的要求。在科学体系的发展之中，对科学理论的深刻、统一和预见力等要求构成明确的价值观，科学体系中新的成熟理论总是在这些价值观所指的方向上高于旧的理论，这就满足了高层建构要求；而随着科学的发展，科学的生长点也按以上价值观所指方向向前推进，这就满足了增长支持的要求。因此科学也是具有进步保证的。

在金观涛所著的《整体的哲学》（四川人民出版社，1987）的“序：理性哲学的理想”中有这样一段话：“总之，今天人类正处

于一个理性主义退潮和日益分裂的世界中。但我认为可悲的并不在于人类尚不能把整个知识统一起来，而在于我们已经满足了这种分裂。人们常常有这样的意见：既然世界是多元的，真理也必然是多元的，因而分裂是正常的，对世界统一的解释也许只是一个古老的梦幻。这种观点混淆了问题。科学上的多元论和探求世界的统一性是认识之剑的两个不同的刃，缺少了任何一个，我们都将变得肤浅而易于满足。彻底的多元主义本身也应该容忍反对多元主义。历史似乎表明，人类在理性科学上的重大进步，几乎都是因为发现了世界看来是相互独立的各方面之间的联系；而最初的起点都是对已经分裂的既成现实的不满。

多元主义之所以有合理的成分，关键在于它主张科学的宽容精神。在多元论的背后隐藏着一个统一，那就是对进步与创新的渴求。哲学家们应该永远充满着无畏的怀疑精神。当一元论居于主导地位时，他应该去怀疑是否已出现了一种迷信；当多元主义思潮滚滚而来时，他同样应该怀疑表面上的多极性或许只是一种浅薄的时髦。思想深刻的哲学家永远在寻找世界深刻的统一性。正如波尔所说，浅薄真理的反面是谬误，但深刻真理的反面照样是真理。我认为这恰当地说明了多元论和一元论之间的关系。”

从哲学的高度来看，软硬结构模型可以把多元论和一元论纳入一个统一的体系之内。从非本体论的认识论角度来看，一元论代表一种追求统一性的价值观，这种价值观的存在理由不是基于“世界是统一的”这种形而上学的信仰，而是由于统一性符合人类认识所依据的基本价值观。多元论则代表一种宽容态度，在确定价值观的指引下，多种的对立理论（变异）的存在有利于我们通过选择得到更好的理论。因此，在理论发展的软硬结构体系之中，一元论表现了硬结构的要求，而多元论则表现软结构的要求，二者缺一不可，失去任何一方都会造成理论发展的停滞。金观涛的说法表现出对于一元论和多元论各自的重要性的深刻洞察，但是

离开了软硬结构模型，关于二者之间的关系我们只能得到波尔那种参禅式的说明。

3.5 生物的进化与进步

对于进化这一概念的最大的误解，就是把进化混同于进步。这种进步的宇宙观可以上溯到亚里士多德的“目的因”，即认为世界上存在着一种发展的目标，万事万物都向着自我完善的或是由低级向高级的方向发展。进化论的先驱者拉马克，在哲学上也认为生命的低级形态在自然条件下由无生物反复产生，“由造物主传来的力”即指向复杂性的天生倾向使这些生物必然地向着复杂和完善的方方向进步。现代科学已经完全抛弃了目的论，这并不是由于对机械论的形而上学的偏爱，而是由于从预见力的要求来看，目的论的解释通常是明显劣于机械论的。如果我们把宇宙看成是必然进步的，那么热力学第二定律这种明显倒退的理论在宇宙中就失去了容身之地。对于相信万事万物的存在都具有某种目的的人来说，掌握非拟人化的纯机械论的进化机制似乎很困难，因此他们总是从进化论中读出一些达尔文和现代进化生物学家都没有承认的东西，在进化中看出进步的意义来。实际上进化的概念并不天然地包含进步意义，既不包含某种目标，也不暗示某种方向。特别是与某些人的常识相反，进化不能被解释为向着人类出现的方向前进的。达尔文对于进化与进步的区别有过清醒的认识，作为自戒，他在笔记本上写下了这样一段话：“关于不同类型的生物不应该去谈论高级和低级。”但是他自己也未必能够总是遵循这一原则。把进化看成进步的另一个思想根源是中世纪哲学的完全原则的遗产之一——“自然主义谬误”。即自然存在的都是善的，自然法则不仅是自然的秩序，也是伦理学的原则，把存在换成应该存在得到宇宙的保证。按照这种观点，进化被解释为进步。所谓社

会达尔文主义者为弱肉强食、无约束竞争，种族灭绝和帝国主义找到了理论根据。不自然就是恶的思想也成为反对避孕、堕胎、同性恋的道德依据。直到现在，目的论的影响并未完全清除干净，在不少拥护进化论的人心目中还存在着宇宙从基本粒子向着轻原子、重元子、分子、高分子、有机物、生命、意识、超意识方向进化的“固有进步倾向”的图景。但是严肃的进化生物学者总是注意区分进化与进步这两个含意不同的概念，认为进化并不必然带来进步，有意识地避免改头换面的目的论的侵入。

以上所说的是基于达尔文进化论的常识。对于广义进化论来说，我们上一节中已经触犯了达尔文进化论的禁忌，给出“进步保证”这个说法。这是由于我们的哲学背景是基于评价的善理，不可避免地会染上伦理色彩。追求自然科学与社会科学之间的深层的统一，是科学的深刻与统一的价值观的体现。只是在追求这种统一性时，要注意防止形形色色的目的论和自然主义的翻版。运用机械论即可防止目的论，注意到评价的相对性，就可以避免自然主义。本节的主要任务是利用软硬结构模型对于生物进化中的进步倾向给出一个机械论的说明，也就是讨论生物进化过程中的“进步保证”，或者说讨论进化的方向性。为了讨论进化的方向，首先需要明确一个稳定的价值取向，以便决定进化到底是朝着哪个方向前进的。一个最自然的想法是根据适者生存的原则，把适应度的提高定义为进化的方向。由于生物的进化不象科学的进化那样受到人类有意识的参与，所以在这里要特别注意避免目的论，我们不能认为宇宙中存在着促使生物走向适应环境的固有倾向，只能根据变异选择的机制使得生物群体中适应的个体数增加来说明这种表面倾向的立足于机械论的机制。我们所说的“生物向适应的方向进化”只是上述机制的一种简化表达。就象我们都知道哥白尼的日心说而仍然说“太阳升起来了。”而不说“地球转过去了。”一样，是一种接近直观的简化表达。在第二章的哲学部分我们建

立了非本体论的认识论和逆镜模型，严格地说，我们不应该谈论任何客观事物，而应该说认识主体中关于客观事物的模型，但是在本书其他地方仍然使用读者习惯的本体论语言作为一种简化表达，否则这本书就很难读懂了。

适应度具有严密的定义，但是把它作为进化的方向并不理想。原因是适应度的指向从全局和大时间尺度来看是很不稳定的，这是由于适应的相对性及条件依存性所引起的。生物进化中的适应是非常局部的和近视的，在某种环境下的适应通常意味着对其他环境的不适应，眼前的适应往往孕育着将来的不适应。这就象布朗运动一样，瞬间看一个粒子是有确定方向的，但是从全局和长时间来看则方向就变得不确定了。所以我们很难说人和细菌相比哪个适应度大，这完全取决于环境条件。因此把提高适应度作为生物进化的方向是不适当的。

纵观整个的进化史很容易产生一种直观的印象，即生物是从低级向高级进化的，高级的生物产生于低级生物之后。因此从低级向高级的进化似乎可以看成进化的方向。但是“低级”和“高级”这样的概念很不清晰，带有太多的恣意性和主观色彩。在这里我们用确定性和复杂性来代替“高级”这个概念。确定性和复杂性都是比较清晰的概念，确定性表示某种不变性或稳定性，复杂性则表示生物组织作为系统的元素（例如细胞数量）增加，细胞种类增加，组织层次增加，系统各部分之间的关系增加和确定，这些都是可度量、可比较的概念。我们把复杂性和确定性作为生物进化的方向，并不意味着任何一种简单的生物都将进化为复杂的生物，并不意味着现代的爬虫类终将进化为哺乳类或鸟类。这种方向性只是表明一种不可逆性，即生物由简单到复杂的进化是可能的，而由复杂向简单“退化”则可能性要小得多，或者说是几乎不可能的。生物学中有一个由归纳得到的法则：当环境的进化压力朝着相反方向作用时，生物的进化不会原路退回。在进化中

所获得的性状经过的时间越长，越不容易失去。例如哺乳类的祖先是鱼类，但是当哺乳类再度返回水中生活时，不会退化成用鳃呼吸，而是象鲸那样，仍用肺呼吸。这就是不可逆性的一种表现。

用软硬结构模型可以解释这种不可逆性的机制。这种不可逆性是在进化过程中基因体系中的软结构对硬结构的建构(硬化)所造成的。对于这种建构我们可以举出两种硬化机制，即横向的“互适应”和纵向的“覆盖”。“互适应”是指这样一种机制：如一种生物通过偶发的突然变异获得了性状 *A*，那么 *A* 只要对生物危害不大就会被保存下来，也可能通过偶然的变异失去，此时性状 *A* 属于软结构。但是如果其后生物又偶然获得了性状 *B*，*A* 与 *B* 之间存在互适应关系，即性状 *A* 与性状 *B* 同时存在时能产生对生物有利的效果，而缺失任何一方都会降低生物的适应度。那性状 *A* 与性状 *B* 之间的互适应关系使得这两种性状都被硬化为基因体系中不可失去的成分，即硬结构。“覆盖”机制则着重于多细胞生物的发育过程。多细胞生物从一个生殖细胞发育到个体的过程是一个自我表述的过程，初始的单细胞可以看成是一台机器，核中的 *DNA* 可以看成是一份包含全制造过程各工序的指令集。第一台机器按最初读出的指令制造第二台机器，第二台机器又读出次段指令去完成后面的工序，直到个体完成。制造过程早期所发生的指令差错可能不影响个体完成，但早期的制造结果被后面工序使用的概率大得多，早期生效的指令发生差错时很容易使后续工序失败，个体无法发育完成。这种强选择压力使得早期指令硬化，成为不可变更的部分，形成硬结构。

以下我们用一些实例来说明上述的抽象机制。对于进化过程中的变化，我们可以把它分为“发明”与“调整”两种形式。这两个概念在某种意义上类似于“突变”与“渐变”或是“质变”与“量变”。发明与调整是相对而言的，调整的积累可能导致发明。发明往往引起新的调整。地形图模型和博弈论模型是处理调整上

的良好工具。发明则缺乏有效的数学模型，突变论可以算是一种用数学方法处理发明的尝试，然而大多数发明是难以量化的。软硬结构模型则主要讨论在大时间尺度上各种发明与调整的相互关系，特别是发明与调整结果的积累过程，小发明积累成大发明的过程。对于地形图模型和博弈论模型来说，发明是提供新的地形图或新博弈规则的过程。而调整是模型中的参数优化过程。

第一个例子是互适应硬化的例子。分子遗传学的研究表明，细胞核与线粒体在遗传上是完全独立的，这暗示着具有复杂构造的细胞是由两种不同的生物结合而成的。我们可以想象一下这种结合的过程。在原始还原性大气下的有机汤海洋中，存在着厌氧性原始真核生物，由于光合作用细菌（原始叶绿体）所发明的光合作用，使原始还原性大气逐步变为氧化性环境。此后由于利用光能反应获得能量的发明导致好氧性细菌（原始线粒体）的产生。在这种条件下，偶然侵入原始真核生物的原始线粒体获得了比有更为稳定的营养供应，而厌氧性原始真核生物又由此获得在不氧化环境下生存的能力。双方的互利使这种合作得以维持，逐步向互适应的方向发展，当有机汤被吃光，氧化环境占主导地位时，这种合作就成为牢不可破的关系，钻出细胞的线粒体和失去线粒体的细胞都无法生存，结合就转化为硬结构了。当然这种硬结构的建立要经过漫长曲折的过程，可以想象在一开始侵入原始真核生物的线粒体可能按寄生者的方式活动，即吃尽宿主而最大限度地繁殖自己。但是由于线粒体自身缺乏运动能力，就象一切依靠宿主的运动才能传播的寄生者一样，受到降低毒性的进化压力，即过快的杀死宿主意味着自身的适应度降低。这个原理甚至对于计算机病毒也是适用的，如果一种病毒一进入一台计算机就无限制地自我增殖，或用其他方法立即破坏计算机的正常功能，那么这种病毒就不可能得到传染给另一台计算机的机会。这种降低毒性的进化压力使得线粒体获得控制自身在细胞中密度的方法，避免

过度增殖，而与细胞建立合作关系的线粒体又能通过增加宿主的适应度来提高自身的适应度，结果由宿主增殖得到的适应度增益远大于传染的利益，进化就使线粒体放弃传染方式而成为宿主细胞的组成部分，通过互适应固定下来，成为硬结构。这也就是在线粒体侵入细胞这一发明所引起的调整使发明硬化的过程。原始叶绿体侵入细胞而形成植物细胞的过程可能也有类似的经历。

作为第二个例子可以观察一下动物胚胎的初期发育过程。这个过程示于图 3.6。图 3.6a 是一个生殖细胞，例如受精卵，先经

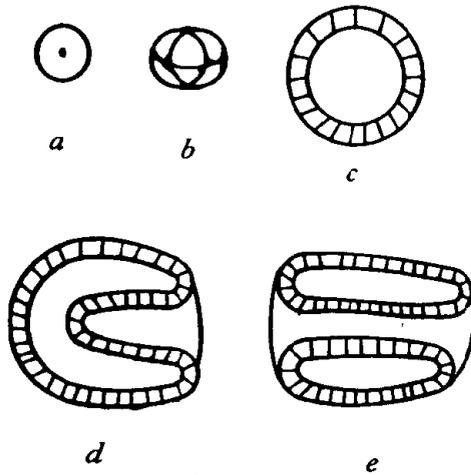


图 3.6

过一变二，二变四，四变八的分裂过程 *b*，然后发育成空心球状 *c*，进而空心球一处产生凹陷，发育成 *d*，*d* 的凹陷底部与外层接合并贯通发育成 *e* 的双层管状。要进行这样的初期发育过程，细胞分裂必须遵循一些由基因规定的原则。从图 3.6 的过程我们可以推断出一些基本的组织原则，例如连接原则，即分裂出的细胞不能自由散开，必须连在一起，否则多细胞生物是产生不了的。还有单层原则，即细胞总是排成单层。再就是近距原则，即各层不能相距过远，避免形成过大的空腔。单层原则导致空心球的形成，近距

原则可以导致当空心球达到一定大小时产生凹陷。如果生物没有在进化中发明这些原则，多细胞动物就不可能产生。从 *d* 到 *e* 经历了一个重大的发明，胚胎的拓扑结构发生了变化，使动物有了嘴和肛门的区别。这个发明的有无使动物分为腔肠动物门和“通肠动物门”（生物学中没有这个名词，这里是为了说明的方便杜撰的）两大类。在“通肠动物”发明的初期，并不显示出太大的优越性，可能适应度还会下降，因为腔肠动物可以把食物在消化道中保存较长的时间，而“通肠动物”吃了东西可能很快会漏掉，但是这一发明产生了新的进化压力，制造了新的地形图和博弈。首先肛门可以象嘴一样成为可开闭的以防止漏出，其次消化道可以变得细长以增加消化和吸收的表面积。消化道变长与肛门和嘴的分化（专业化）是互适应的，长的消化道要求食物单方向移送，肛门与嘴必须分工。经过这样的调整，发明“通肠动物”的意义就开始明确了，动物对食物的处理过程由腔肠动物的“批处理”的间歇工作方式变成“通肠动物”的“流水线”连续处理的方式，由肛门的发明和消化道的调整导致了“流水线”的发明。流水线消化方式造就了消化的前后关系，为消化与吸收的分化准备了条件。在腔肠动物门的消化道里，消化道内壁细胞必须同时具备分泌消化液和吸收营养两种功能，而在流水线中，由于有了前后顺序，先消化后吸收能够明显地提高效率，前部的消化道逐步失去吸收功能，后方的消化道逐步失去分泌功能，双方的互适应导致了胃和肠的分工。这是一个由调整积累成的发明。纵观整个消化道的进化过程，可以看出一个发明怎样引起调整，调整又导致下一个发明，而后续的调整和发明怎样使前面的发明硬化。对于原始的“通肠动物”来说，通肠方式只相当于多了一个嘴甚至漏洞的腔肠动物，此时偶然的变异还可以使通肠退成腔肠，但是当嘴和肛门分化，流水线产生，特别是胃肠分化之后，通肠退化成腔肠就完全不可能了。如同中国人说的骂人话，作事太缺德，生孩子没屁

眼儿，就是指一种由通肠向腔肠的退化，在不会作外科手术的动物世界里，这种变异个体是不可能成活的，一定会被淘汰。这就是后面的发明对前面发明的覆盖硬化作用。肠和胃的分化则是互适应硬化，这些不可逆过程实现了软结构对硬结构的建构，而发明导致调整、调整积累成发明的过程则是硬结构不断促使软结构增加，对软结构实现支持作用。“通肠动物”的发明是一个关键性的发明，腔肠动物门在其后的繁荣远不如通肠动物，只包含象海蜇、水螅这种低等动物，而通肠动物则产生了众多的动物门。但是这个关键性的发明本身并不太困难，只要凹陷的内层伸长一点贴上外层，再加上单层原则的作用使内外层贴在一起的部分的细胞坏死，就成了通肠。

互适应硬化和覆盖硬化都是不可逆的。生物体各部分之间的相互适应不会调转方向向不适应前进，胃和肠一旦分化就不可能再向通用方向进化，分工合作结构一旦生成就不会解体，这意味着生物只能朝着由简单向复杂方向进化。从发育过程的覆盖硬化来看，逆方向的进化即沿原进化方向逆向退回会遇到双重的障碍，一个是调整和互适应的不可逆性，另一个是进化过程中变异的随机性和选择的方向性。即使一个进化过程不包含内部互适应，要想让进化原路退回也必须满足两个条件，即按原来进化过程完全相反的顺序产生变异，同时又按当时相反的顺序受到进化压力。按照统计学的常识，这两个条件都是不可能满足的，这相当于掷多次骰子之后，又要求按原来相反的顺序再掷出同样点子一样。

维纳在他的名著《控制论》(Cybernetics)一书的开头提到两种时间观。一种是基于牛顿力学的时间，是可逆的。如果用摄影机拍下行星的运行，然后把影片倒退放映，影片中的世界依然符合牛顿定律。另一种时间是基于热力学第二定律的时间，是不可逆的。我们可以把冰放在热水里得到一杯凉水，但不能期待一杯凉水自动变成一杯热水同时结出一块冰。这种不可逆性是由熵的

只增不减保证的，可以称之为“熵垒”。“熵垒”是由统计学支持的。我们所描述的基于软硬结构模型的时间可以说是第三种时间，它也是不可逆的，这种不可逆对应着局部的熵减少，信息量增加，系统向高度有序的方向发展。保证这种不可逆性的是“硬化壁垒”。从上述掷骰子的说明可以看出，“硬化壁垒”也是由统计学支持的。与“熵垒”相比，“硬化壁垒”具有更积极的意义。

“硬化壁垒”所保证的不可逆性使得生物进化从大时间尺度上看只能由低级向高级的方向前进或者停滞，而不可能倒退。这种进步表现在发育过程的工序越来越多，成体结构越来越复杂，确定性越来越高。例如比较古老的软古鱼类鲨鱼的牙齿掉了还能再长，爬虫类在同种中牙齿的个数也会有差异，而哺乳类则在不同的属之间也有时牙齿数是一样的。骨格结构的确定性也可以看出大致相同的趋势。这是由于覆盖硬化造成的。另外在进化过程中，四肢的反复进化、退化时有发生，而内脏组织则非常稳定，一旦生成就不再消失，这是由于四肢主要是实现与外部环境的适应，而内脏之间则是分工合作的内部互适应关系，一旦形成就难以解体，不经覆盖就自行硬化。

在进化与发育之间存在着称为“重演律”的关系。这种关系的早期表达形式是“个体发育过程重复系统进化过程”。当然对于重演律不能作如此简单的理解，不能认为胚胎发育的某个阶段就是一种曾经存在过的生物成体。如果运用软硬结构模型，可以对重演律作出更为深刻的解释。重演律所表达的一个重要的含意是，多细胞生物在进化中所获得的首先是制造成体的工序（发育过程），其次才能谈到这些成体的性状和性状对适应度的影响。如果一道工序依赖于另一道工序的制造成果，这两种工序之间就有了确定的先后关系，而这种先后关系在进化与发育的过程中当然是一致的。在进化过程中，前工序还没有时后工序不可能被发明；在发育过程中，前工序未完成则后工序无法开始。而且发育过程中

的制造方法（工序）除了进化过程中的发明之外不可能有别的来源。从软硬结构模型的角度来看，发明显然不是任意的，而是受到很大约束的。已经硬化的发明不可能被取消，只能添加新的发明。例如在胃肠分化之后不可能再合二为一，只可能添加胆囊、肝脏之类；脊椎动物由脊索动物进化而来，在脊椎动物发育过程之中仍保存着脊索的阶段，因为神经系统的早期发育要依靠脊索的引导，这种覆盖硬化使脊索成为脊椎动物发育的必经之路。因此发明只能在软结构中产生。由此我们就可以解释为什么在生物进化过程中表现出发散与收敛的交替。由于发明需要很多条件和具有一定的随机性，因此在进化过程中大的发明不是随时能产生的，而在一个革命性的发明之后，往往导致各种调整，产生丰富的软结构，各种小发明应运而生。在此阶段，进化呈发散形式，大量的新种产生出来。而随着进化的推进，在新的大发明出现之前，前一个大发明所引发的小发明逐渐硬化，生物或在稳定凹里变化不大，或者由于硬化引起的对进化压力反应能力下降而导致灭绝。此时进化呈收敛形式，直到下一个大发明到来，才能引发下一次发散。因此动物的几乎所有的门，都在进化的早期就已经形成，“通肠动物”的发明可能是这次发散的起因。这种大发明在科学的进化之中对应着库恩所说的“科学革命”。对于现存的复杂动物来说，前面的发明已经硬化，因此不可能从中产生出新的动物门。

纵观上述的进化机制，我们可以对进化重演律作出更为准确的描述。生物发育的各个阶段并不是进化过程中各个时期动物成体的形态，而是各时期中被硬化的发育工序的积累。几乎可以断定，发育过程中后期形成的性状，一定是进化后期发明的产物。生物胚胎发育所需的各个工序即各阶段的制造方法只能来源于进化过程中的相同顺序的发明，这是进化重演律的理论根据。

为了以广义进化论的观点对于生物具有软硬结构的进化过程给出一个定量的说明，我们需要建立“硬信息量”这一概念。在

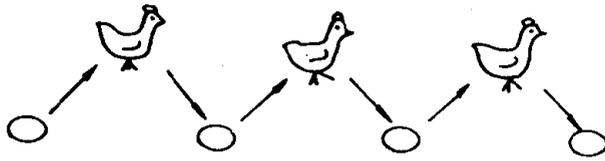
一种生物所拥有的大量遗传基因之中，各个基因的重要性是不同的，有些基因发生变化对生物的生存毫无影响，也有的基因一旦变化就会引起致命的效果，前者我们称之为软基因，后者称之为硬基因。当然也有不少基因是半软半硬的，即发生变化时生物的成活率会受到影响，但并不是一定致命的，这种基因可以折合为不到一个的硬基因。对于已经硬化的发育工序起指令作用的基因，通常是硬基因。在同种的生物个体之中，硬基因的个体差很小，而软基因可以有较大的个体差异。硬基因由于受到严格的选择压力，可能表现出长期稳定不变，而软基因则在发生变异时不被除掉，可以呈现与时间成正比的变化。因此测定进化时间的“分子时钟”通常使用软基因，测定个体差异或亲子关系所用的“基因指纹”也是软基因。对于一个生物种来说，全部硬基因所包含的信息量，称为该物种的硬信息量。硬信息量可以粗略地理解为硬基因的个数。但是由于我们目前对于复杂生物的基因地图还没有掌握，对于各个基因在发育和生存、繁殖中所起的作用还知之不多，所以由基因的组成直接计算硬信息量目前还不可能。从当前的科学技术水平出发，我们可以找到间接测定硬信息量的方法，使得关于硬信息量的假说成为可验证或可证伪的。由于硬基因的破坏可以导致生物的适应度下降，因此通过可控的提高突然变异率，同时测定生物适应度变化的程度，就可以推算出硬信息量。基因突变引起的影晌通常是坏的，对生物有利的突变只占其中的一亿分之一，可以忽略不计，因此适应度的下降可以作为破坏的标准。促进突然变异率有很多物理的和化学的方法，例如放射线或化学药品。控制放射线或化学药品的剂量就能控制基因突变率。调节突变率使得被试生物的适应度降为二分之一，根据突变率和适应度的降低，运用统计学的方法就可以计算出硬信息量。即使不进行计算也可以作出相对比较，在适应度下降程度相同的情况下，相应的突变率越低表示该种生物的硬信息量越大。这个实验有待生物学家来

完成，目前我们只能通过从硬信息量假说所能作出的一些推论与生物学的已知事实是否相符来对这个假说作出初步评价。

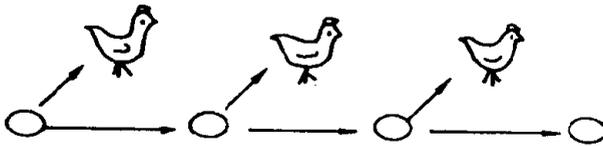
生物在由低级向高级的进化（进步）过程中，硬信息量只增不减。随着硬信息量的增加，生物对于突然变异的耐受能力必然下降，例如对于硬信息量为 10 的生物来说， $1/1000$ 的基因突变率只使其适应度下降 1%，而对于硬信息量为 500 的生物来说，每代 $1/1000$ 的基因突变率会使适应度下降一半，相同的突变率能使硬信息量为 2000 的动物的后代几乎不能成活。因此随着生物硬信息量的增加，其突然变异率必需降低。这个推论与生物学中的实际观察是相符的。要作到这一点生物必须具备控制自身基因突变率的手段。这个手段也确实存在的，例如有的基因本身可以促进其他基因的突变，而一些基因修复的机制又可以抑制基因突变。高级动物与植物相比，发育到成体所需要的工序要多得多，其硬信息量也大得多。其复杂度也要大得多，组织、系统远为复杂，细胞种类也多。动物比植物的确定性也大得多：从成体的大小来看，动物是一定的，而植物不确定；从外形来看，树木被砍去一些枝条不会影响生存，而动物缺了一个肢体也将导致生存能力明显下降；从组织特异性来看，不同种的植物之间可以嫁接，而在动物中同种的个体之间的器官移植也总是遇到异体排斥的困难。由于硬信息量大，动物在发育过程中尽量避免遗传基因突变的可能性，其中一个重要措施就是减少生殖细胞所经历的细胞分裂次数。动物胚胎发育过程在经过十几次细胞分裂之后，生殖细胞就被保存下来，待其他部分发育成熟之后，这个生殖细胞才被起用来作为制造生殖细胞的母细胞。而植物由于其硬信息量小，完全不必设法防止突变，从一颗种子到树顶上的果实，经历的细胞分裂次数多得难以计算。在一颗树上甚至会在某个枝上长出与其他枝明显不同颜色和形状的树叶，突变可以随时随地地发生。从某种角度看，与动物相比，植物体内的组织要松散得多，很象是一枝珊瑚，

珊瑚的形态很象树也说明这种共性，当然植物体内的各细胞比珊瑚虫之间的分工合作关系要密切得多。由于植物与动物之间硬信息量的差异，对于植物来说可以收到一定效果的放射育种对于动物则难以使用。很小的放射剂量不足以诱发足够的变异，由于动物的硬信息量大，稍增加放射剂量又会使动物完全失去生殖能力。

“先有鸡还是先有蛋？”是一个古老的问题，人们常常用它来比喻一种无限退行或是无法解决的问题。但是我们如果理解了进化重演律及其原因，用前面分析过的遗传和发育模型可以对这个问题给出一个明确的回答。这个回答就是“先有蛋”。通常我们所想象的鸡与蛋的关系如图 3.7a 所示，即鸡生蛋、蛋生鸡的过程。但是如果仔细研究鸡的发育过程，就会发现这个模型并不准确。在鸡还没有发育完全，即蛋只经过十几次分裂之后，下一代的蛋已经被保存起来，不再参与鸡的发育了。从这一点上看，蛋是由前一个蛋直接生的，而不是鸡生的。从进化重演律的角度来看，发育是进化在硬结构上的重演，具有严格的不可逆性，在进化过程中，后工序一旦使前工序硬化，其后的发育变化不可能改变前工序；在发育过程中后续的工序由于时间的不可逆性不能改变前工序。因此蛋中的基因指令完全决定鸡的形态，而鸡对于改变蛋中的基因是无能为力的。从这一点上看，蛋能生鸡，而鸡是不能生蛋的。根据这两点我们可以得到如图 3.7b 那样的模型。在这个模型中，不是鸡生蛋、蛋生鸡，而是蛋生蛋、蛋生鸡，因此从因果关系来看，对“先有鸡还是先有蛋”这个问题的回答当然是“先有蛋”。在这里讨论“先有鸡还是先有蛋”的问题并不是为了反驳拉马克的“用进废退”说，我们所感兴趣的是蛋和鸡的关系对应着认识论中的理论与实践的关系。在这个关系中蛋对应着理论，而鸡则对应着实践。理论可以指导实践，而在实践中不可能自发地产生理论。这里也有一种不可逆性。即从事实中推导出理论的合乎逻辑的手段是不存在的，就如同从鸡的结构不可能推



a



b

图 3.7

导出蛋中的基因一样。理论只能靠前面的理论进化而来，实践只能产生对理论的需求和验证，但不能产生理论本身，就如同鸡的适应度可以决定蛋的存亡却不能改变蛋的基因一样。从软硬结构的角度看，鸡的发育能否成功绝大部分是由蛋壳内的发育过程所决定的，一个理论在接受实践的验证之前已经在大脑中或学术讨论中经受过象一致性、明晰性等多种评价准则的检验。

植物的情况与动物有所不同，鸡不能生蛋，但是核桃树能结核桃。核桃树从核桃发育而成，核桃树顶上的树枝经历了多次的细胞分裂，如果在这个过程中发生了变异，那么这个变异是可以保存在下一代的核桃之中的。如果把树看成一个整体，那么从宏观上看树木是有可能表现出拉马克进化的样子。因此米丘林—李森科学说也不能说完全没有道理，他们的实验结果，主要来自于植物也不是偶然的。但是如果把树看成细胞的有简单分工的松散联合体，以细胞为单位来看问题时，这种拉马克式的进化仍然被还原成孟德尔—摩尔根遗传的积累。因此孟德尔—摩尔根学说比米丘林学说更为深刻。核桃树可以结出核桃说明实践也可以产生

理论，但只能产生肤浅的理论，类似于公鸡哲学中明天还可以吃到米那样的理论。由实践产生小理论的逻辑是存在的，神经网络的学习就是运用这种逻辑的例子，但是单凭这种逻辑不可能直接产生深刻的理论，因为神经网络学习的方法对于复杂理论来说是指数爆炸的。

3.6 有性生殖的意义

人为什么分成男人和女人？关于这个问题古往今来有各种理论和解释。按照基督教的说法，上帝造人时先造了男人后造了女人。这当然算不上一个理论，因为它没有说明上帝为什么这样造。按照中国古代的阴阳五行学说，阴阳是宇宙的普遍规律，因此人也分为阴阳两类，即男人和女人，如同天上有太阳和月亮（太阴）一样。这个理论当然也不能令人满意，它不能说明为什么细菌不分为雄性和雌性。

近代生物学也不断对这个问题作出回答，达尔文就已经对性选择在进化中的作用作过详细的研究。但是长期以来由于理论工具的缺乏，这方面的研究没有太多令人信服的成果。例如在 60 年代，我曾见到过这样一种解释：变异使基因产生缺陷，有性生殖由双方取得基因，其缺陷部位不同，可以互相弥补。这相当于信息理论中利用冗长性来纠正误码的机制。但是这个说法用进化论的眼光来看是没有说服力的。因为无性生殖可以在基因缺陷发生时立刻将带有缺陷基因的个体淘汰掉，使得缺陷基因不能存在，所以用冗长性来纠正误码是没有必要的。直到 70 年代以来博弈论方法与进化生物学相结合，进化论的研究重点由种群向个体，由个体向基因方向发展，博弈论的数学工具与计算机模拟的方法在研究中的推广，使得关于有性生殖的研究取得了重大进展，使我们对于有性生殖的进化过程逐步有了较为清晰的认识，对于很多环

节例如两种性细胞的分化，雌雄个体的产生，性比的进化等都可以用比较严格的数学方法主要是博弈论的方法给出相当严密的论证。

在单细胞或简单的生物中已经有了象细胞融合这样的“准有性生殖”方式，但此时并没有雌雄之分，相互融合的两个细胞是平等的。例如象水绵这样的植物，通常是用细胞分裂的方式繁殖，偶尔也采用“准有性生殖”方式，两条水绵丝并在一起，相对的细胞互相融合，产生新的生殖细胞，然后“并丝”解体使生殖细胞散开，各自长成新的水绵丝。对于水绵来说，生殖细胞就是体细胞，二者没有分化，因而生殖细胞的大小是由体细胞的最佳大小决定的。当生物进一步复杂化之后，由分工的发明产生出专门的生殖细胞，而且在复杂的发育工序完成之前生物还没有独立生存的能力，其能量和物质都要由生殖细胞提供，此时生殖细胞的大小对于生物的适应度就会产生很大影响，这种进化压力使生殖细胞的尺寸通过进化调整到一个最佳值。生殖细胞的尺寸大，含有能量和物质多，其存活率也会提高，但是由于生物所能积累起来用于生殖的能量是有限的，生殖细胞的尺寸加大意味着数量减少，而存活率的提高也是有限的，不会超过 100%，一个存活概率为 1 的生殖细胞不如两个存活概率为 0.6 的生殖细胞的适应度大。所以生殖细胞决不是越大越好，对于一种生物来说存在着一个最佳的生殖细胞尺寸。当生殖细胞由两个细胞融合而成时，按照简单的计算是每个细胞都是最佳尺寸的 $1/2$ 时适应度最大，而且公平合理。但是，生物在进化中不会自己计算最佳值，最佳值是变异与选择的结果，是用“试错法”算出来的。进化游戏中不存在公平原则，进化是自私的基因之间的博弈。我们假定在博弈开始时存在着各种不同大小的细胞，每两个细胞可以融合成一个生殖细胞，各细胞是随机相遇的，就象细胞不能计算自己的最佳尺寸一样，也不能计算与多大的细胞融合最上算，一切都按变异

与选择的原则行事。各种大小不同的细胞数量是不同的，越大的细胞则数量越少，大小与数量成反比。为了说明的方便我们把最佳尺寸看成临界尺寸，即大于等于临界尺寸的生殖细胞生存概率为 1，而小于临界尺寸时生存概率为 0。由于大小与数量成反比，细胞与小细胞相遇的概率大。通过计算可以发现，在各种尺寸中，等于临界尺寸的细胞和极小的细胞适应度最大，而比临界尺寸小但又小得不多的，例如 $1/2$ 临界尺寸的“公平”细胞适应度很低，会被淘汰。为了简化计算我们考虑存在三种尺寸的细胞。一个临界尺寸的，两个 $1/2$ 临界尺寸的，1000 个 $1/1000$ 临界尺寸的。在这个系统中，临界尺寸的细胞无论和哪个细胞结合都可以存活，其生存概率为 1， $1/2$ 临界尺寸的只有与 1 或 $1/2$ 的结合才能存活，其概率为 $2/1002$ ，适应度为 $4/1002$ ，而 $1/1000$ 尺寸的细胞只有遇到尺寸为 1 的细胞才能成活，其概率为 $1/1002$ ，但是由于这种细胞有 1000 个，适应度为 $1000/1002 \approx 1$ 。更复杂的初始条件设定算起来比较麻烦，用计算机模拟的进化结果与上述简化情况是一致的，即细胞（配偶子）的大小必然两极分化，一极达到最佳尺寸，另一极则在不损害其配偶子功能的条件下尽可能小而多，中间状态是不能存在的，“公平”只是一种幻想。配偶子在尺寸博弈中只有两种战略可取，一个是提供足够的资源，成最佳尺寸，一个是以数量取胜，成最小尺寸。前一种战略产生卵，后一种战略产生精子，中间路线是不能存在的，雌雄性配偶子的两极分化具有数学的必然性。

在雌性和雄性配偶子分化之后，作为生物个体，雌雄同体的现象还长期存在，但是由于精子的制造成本比卵低得多，在雄性功能竞争激烈的情况下，一部分个体放弃生产卵而全力投入竞争可以使适应度增加，从而产生了专业的雄性，一旦专业的雄性产生，雌雄同体的个体在雄性功能方面难以与之对抗，雄性功能的价值下降，而放弃雄性功能成为专业的雌性能够提高适应度，从

而产生了专门的雌性个体，完成雌雄异体的分化。因此在向雌雄异体的进化过程中，总是先产生雄性，后产生雌性，这与实际的生物观察结果是相符的。

由于精子的制造成本很低，一个雄性个体可以使很多雌性个体的卵受精，从生物种群的整体利益来看，雌雄性比为5：1比1：1的情况适应度要高得多，但是进化通常是不顾整体利益的。在性比博弈之中，如果种群之中雌性个体多于雄性，则多生雄性个体的基因的适应度会提高，反之也是一样。因此整个群体的成年时性比总是保持在1：1，这是性比博弈的稳定点。

对于有性生殖进化过程中的一些关键环节和很多具体现象，现代进化生物学运用博弈论的方法作出了满意的说明。但是有一个最根本的问题，即有性生殖本身为什么能够存在，至今还没有一个有说服力的理论。问题的关键在于，有性生殖需要一个父亲和一个母亲，而无性生殖只要有一个母亲就够了，这意味着，有性生殖的成本是无性生殖的两倍。从适应度的角度看，有性生殖的适应度只有无性生殖的二分之一。在进化的过程之中，两倍成本将会造成很强的进化压力，使有性生殖被淘汰掉。因此关于有性生殖的进化理论不仅需要举出有性生殖的优越性，而且需要说明这种优越性足以产生2倍以上利益，以抵消掉二倍成本的损失。这就是建立关于有性生殖的基本理论的困难之处。

有性生殖有一个众所周知的优越性，就是通过基因的重新组合，可以产生众多的变异个体。关于这一点，我们需要讨论一个重要的观点，即“多样性的价值”。进化的基本机制是变异和选择，变异越多，选择的余地越大。无论是对环境变化的适应还是生物体内部的发明和调整，都只有在拥有充分的变异时才可能进行。举一个极端的例子，如果在一个生物种群中选择出一个“最佳”的个体，然后用这个个体的完全拷贝来构成整个种群，那么这个全部由最佳个体构成的“最佳种群”却是最不佳的种群。由于失去

了变异，选择也无法进行，当环境变化时，这个种群对于进化压力不可能作出任何反应，只有走向灭绝一条路。这就说明了“多样性的价值”是与眼前的“最佳化”相对立的。在给定的评价准则下只能有一个最佳者，因此“多样化”意味着容忍一大批不佳个体的存在。由于生物不可能预测未来环境的变化，只能靠拥有足够的变异作为进化的准备，以适应无法预知的环境变化。大量的变异以牺牲“最佳”为代价，换取了生物种群的“柔韧性”和“顽健性”，取得长远利益与眼前利益的平衡。

关于有性生殖的优点有很多进化生物学者进行过多方面的研究，其中有从长远利益即集团选择的方向研究的，也有从眼前利益即个体选择的角度研究的。集团选择论者作出的解释是，基因的重组有利于生物种群增加对环境的适应力，无性集团的种灭绝率比较高。设想有一个高频度存在基因 A 和 B 的集团，假定在环境变化时，稀有的变异 a 和 b 相互作用会使适应度提高，在无性生殖种群中，变异 ab 只能经过 aB 或 Ab 的变异再由它们的子孙产生，须要很长的时间，一下子产生 ab 的概率是极小的。而在有性生殖种群中，只要有了 aB 和 Ab ，就能通过基因重组立即产生 ab 。这样有性生殖对于环境变化的适应要快得多。这种说法得到一些观察结果的支持，从分类学上看，类似的生物如果同时存在有性生殖与无性生殖种，那么无性种远不如有性种繁荣，而且无性种大多历史比较短，说明无性种适应能力差，容易灭绝。但是这个理论有很多缺点，在小集团中， aB 与 Ab 同时出现的概率也是很低的，有性生殖种并不显示什么优越性，而且对于有性生殖来说即使出现了 ab ，在下一次的繁殖中这种优越结合还是被打散的。进化通常是不考虑长远利益的，一种性状如果具有长远利益而在眼前有害，那么它在长远利益显示之前就已经被淘汰掉了。因此，理论家们也在寻找有性生殖的眼前利益。一种模型是考虑有一个非常复杂的环境，在空间上分为马赛克状的小环境，那么由重组

而产生的多种基因变异可以比单一的基因更容易适应这种环境。另一种模型是考虑在同一个环境之中有复杂变异的种群由于变异而导致要求不同的资源,可以通过减弱种内竞争而提高适应度。然而这两种模型的缺点是要求的条件太强,因而缺乏普遍意义。提出有性生殖二倍成本问题的威廉·汉米尔顿(William D. Hamilton)用《爱丽丝漫游奇境记》中的红桃皇后赛跑的故事来作为比喻,主张用“红色皇后”(Red Queen)理论,在时间轴上解释有性生殖的优越性。这个理论认为,是寄生者与宿主之间的“军备竞赛”使得宿主不得不采用有性生殖的方法来逃避寄生者的危害。用一个简化的模型来说,寄生者有 A_1A_2 两种对立基因,宿主有 B_1B_2 两种对立基因,持有 B_1 的、宿主只对持有 A_1 的寄生者有抵抗力,而 B_2 只对 A_2 有抵抗力。当 A_1 的寄生者比较多时, B_1 的适应度高,使得 B_1 型宿主增加,从而引起 A_2 的适应度提高,又引起 A_2 型寄生者增加,因此又会引起 B_2 型宿主增加,如此产生周期性的振荡。而在这种振荡条件下,有性生殖的快速重组是有利的。这种景象就如同童话中的“红色皇后”一样,在原地飞快的转圈子而并不前进。这是目前比较有影响的有性生殖理论。这个理论也受到不少观察实例的支持,从分布来看,在高纬度,干燥,小水域等物理环境多变而恶劣的地方,无性生殖种多,而在热带,湿润、寄生等物理环境优越而生态环境复杂,生物种间竞争激烈的条件下有性生殖种占优势。

上述关于有性生殖何以能存在的现有理论有一些共同的缺陷。第一个缺陷是缺乏普遍性,就连红色皇后理论也要求相当苛刻的条件。在普遍的吃与被吃的竞争之中,未必都能与基因的变异扯上关系,很难想象持有某一种基因的狮子不吃持有某一种基因的羚羊。第二个缺陷是,这些理论都能指出有性生殖在某种条件下的优越性,但是都不能说明这些优越性何以能达到二倍以上的适应度,足以抵消二倍生殖成本的支出。第三个缺陷是,这些

理论只能作小时间尺度的说明，即在既存在有性生殖种又存在无性生殖种的近缘种之间，是什么因素使其走向有性或无性。在大时间尺度上，即对于为什么高级生物比低级生物更强地倾向于有性生殖这个问题上，这些理论都是无能为力的。在一注臭水之中生存的各种微生物之间，同样存在激烈的竞争和军备竞赛，为什么细菌几乎不采用有性生殖的方式，尽管所有的受精过程都是单细胞的行为，单细胞生物也不是没有发明这个方法。反之鸟类和哺乳类这种高级动物不管生活环境如何，几乎全部是有性生殖。这又是为什么呢？

运用软硬结构模型及多样化的价值和硬信息量的概念，我们可以对有性生殖的意义给出一个更具普遍性的说明。在上一节中对于生物进化的方向性的分析表明，生物进化是一个从低级到高级的不可逆的过程。在这个过程中，生物基因的硬信息量是只增不减的。随着硬信息量的增加，生物不得不减少突然变异率，以维持最低限度的生殖成活率即适应度。但是对于生物来说，又必须维持一定多样性，否则生物无法对进化压力作出反应。对于无性生殖的生物来说，多样性只能依靠突然变异来实现，生物的基因突变率面临着两个相反的要求，多样化要求突变率大，而成活率要求突变率小。对于硬信息量很小的简单生物来说，要让基因突变率同时满足两方面的要求是可能的，但是当硬信息量增大，要求进一步降低突变率时二者之间就会产生不可调合的对立，折中方案将不复存在，此时有性生殖将必然取代无性生殖。有性生殖可以通过基因的重组来实现多样性，而不必增加突变率。基因突变是随机发生的，因此形成对软基因和硬基因的“无差别攻击”。有性生殖的基因重组在同位基因的交流之中，由于硬基因在所有个体中都相同，交换并不引起变化，而软基因各有不同，利用重组所产生的“组合爆炸”可以千变万化。据估计同一对夫妇可能生出的子女有几百万种不同的组合，所以除了一卵性双生子之外，

几乎不存在兄弟或姐妹完全相同的现象也就不足为怪了。

从利益与成本的角度来看，有性生殖与无性生殖比较，其相对成本为二倍，利益则随生物硬信息量增加而增加。对于硬信息量很小的简单生物来说，很高的突变率可以使突然变异所造成的多样性满足实现对进化压力发生反应的要求，此时有性生殖几乎没有利益。随着进化过程中硬信息量的增加，突变率被迫降低，为满足多样性要求基因重组引起的多样性所占的比例越来越大，有性生殖的利益也就越来越大。因此必然地存在一个象经济学中那样的盈亏临界点，在这一点有性生殖的利益达到二倍，也就是抵消了二倍成本的损失。到达盈亏临界点时的硬信息量我们称之为临界硬信息量。硬信息量远小于临界硬信息量的生物例如单细胞生物普遍采取无性生殖的方式，硬信息量远大于临界点的生物，例如鸟类、哺乳类则只能采用有性生殖的方式。硬信息量处于临界点附近的生物，例如极简单的动物和多数的植物，则可以有两种生殖方式，视环境对多样化的要求不同而采取不同生殖手段。在这一阶段，生物之间的“军备竞赛”当然是影响多样化要求的重要因素，象复杂的环境空间等其他因素也都对多样化要求有影响。因此基于硬信息量的解释与现有的其他关于有性生殖的理论并不矛盾，而是互相补充的，其他理论侧重于小时间尺度的横向比较，而基于硬信息量的理论侧重于大时间尺度的纵向比较。

从软硬结构模型的角度看，有性生殖的意义在于实现了软硬结构的分离。对于某种生物的一组基因来说，其中的硬基因相当于硬结构，软基因则相当于基因组的软结构。有性生殖可以在降低突变率以减少硬基因破坏的同时通过重组实现软基因的多样性，使软硬结构各司其职，各得其所。当然，在有性生殖种中，软基因的变异从根本上来说仍然是基因突变的产物，但是有性生殖对于变异起到保护作用，使其不会象无性生殖那样容易失去。我们可以考虑两个无性生殖的细菌落到一碗肉汤里繁殖起来的情

况，如果不考虑突然变异，那么这一碗肉汤里所有的细菌都只可能有两种不同的遗传性状。只有通过频繁的突变，才可能产生遗传变异。如果是一对鸟飞到一个孤岛上，情况就大不相同了，这两只鸟可以有4套不同的基因，而在交配所生的下一代里可以产生大量不同的遗传组合。在强选择压力下生物个体数变少时，无性生殖种的多样性随个体减少而成正比的减少，有性生殖种却可以保护其多样性，减少软结构的损失。任何维持多样性以争取长远利益的措施都要以牺牲眼前的最佳为代价，有性生殖也不例外。有性生殖有助于保护一些有害的隐性基因。无性生殖时有害的基因很容易被自然选择所淘汰，而在有性生殖时隐性的有害基因当其在种群中所占比例极小时是不容易被清除的。

硬信息量远远大于临界点的高级动物在漫长的进化过程中其有性生殖方式本身也可能被硬化，成为硬结构的一部分，此时就不再会出现单为生殖的变种。在采用有性生殖的方式之后，会进化出一些雄性对下一代进行投资的性状，使得有性生殖成本下降，不再是二倍。这也是使有性生殖硬化的一个因素。由于植物的硬信息量小，即使是有性生殖的植物也可以进行无性的营养生殖，从植物中取出少量可以分裂的细胞，可以培养出完整的植物。曾经有人想象过用人身上的上皮细胞来“复制天才”。由于人的硬信息量很大，这种方法违反了“鸡不能生蛋”的原则，所以不可能成功。从蛋生蛋的原理出发，“复制天才”只能在受精卵早期分裂时用和一卵性双生子相同的方式制造出来，但是那时我们还无法知道他是不是天才。这就是“复制天才”难以成功的理由。

由有性生殖产生的一个重要的课题是性选择。最先注意到雌性的性选择在自然界中的重要性的是达尔文，他主要是从对极乐鸟及其他一些鸟类的研究中得到了这个推论。在“一夫多妻制”的动物之中，雌性的选择会对雄性造成很强的进化压力，迫使雄性

向着雌性喜好的方向进化。达尔文关于性选择的见解一度受到怀疑和冷落，但现在已经得到大量的实验支持，在理论上也有很大发展。在通常的有性生殖的动物种之中，雌性对于后代的投资远大于雄性，因此对于雌性来说，无效的或是获得低劣基因的交配将使其适应度大大下降，这一点促使雌性更严格地选择配偶。但是对于雄性来说，由于其对于后代的投资小，一次失败的交配在总适应度上引起的损失不大，因此雄性对雌性的选择性不强。对于雌雄个体向后代投资大体相等的“一夫一妻制”的动物来说，性选择相对减弱，性的二型性不太明显。当雄性对后代投资大于雌性时，例如某种鸟类，雌性只管产卵，孵化和育雏全部由雄性承担，形成“一妻多夫制”，此时雄性对雌性的选择变强，雌鸟比雄鸟长得更漂亮。从选择者的利益出发，性选择应该集中于对提高适应度起关键作用的性状。对于雄性通过争斗决定胜负而胜者拥有多妻的物种来说，雌性选择争斗中的胜者对于提高适应度起关键作用。而对于雌雄投资相近的“一夫一妻制”动物来说，选择投资能力强的雄性是有利的，例如有些鸟类的雌鸟在择偶时重点考察雄鸟的觅食能力。一般说来，性选择与自然选择的方向应该是一致的，斗争中的胜者意味着体力和机敏性较好，其生存能力和种间竞争能力也是较强的。从这一点上看，把性选择从自然选择中分出来单独讨论的意义不太大，性选择仅仅是自然选择的一个二级方式而已。从这里可以看出基本价值观对于次级价值观的决定作用。我们更感兴趣的是另一种现象，即性选择走向降低适应度的可能性。例如雌孔雀总是倾向于选择长尾的雄孔雀，而长尾巴显然妨碍雄孔雀的行动能力，可以说是降低适应度的，这样一种“反适应”的性状是怎样进化而来的呢？

费舍 (Fisher) 在 1930 年所写的《自然选择的进化理论》(*The Genetical Theory of Natural Selection*) 一书中，在讨论性比问题的同时，讨论了性选择的“逃离”(*Run away*) 效应。当时博弈论

这个用语还没出现，但是性选择可以说是用博弈论的观点最早讨论的进化论现象之一。其后 Lande (1981b) 和 Kirkpatrick (1982) 分别用多基因和二基因模型对性选择进行更精确的研究也得到了类似的结果。费舍的论述大致是这样的：假如在某个时期，动物种群中绝大多数的雌性个体都乐于选择某种性状取极大值的雄性个体作为交配对象，例如选择尾巴最长的雄鸟。如果这个动物种是一夫多妻制的，则雄性尾巴越长适应度越高。如果尾巴的长短是遗传的性状，那么选择长尾雄鸟的雌鸟生出的儿子尾巴也比较长，其结果是孙子孙女的个数增加。由于这种机制，无论是对于雄性的长尾还是对于雌性喜好长尾都产生很强的进化压力，这种进化将持续到过长的尾巴受到自然选择的制止为止。动物种群的这种进化结果，即由于尾巴过长而明显降低生存概率的状态，可以看成非对称博弈的 ESS。在稳定点上，雄性有了短尾则得不到配偶，适应度降低；而雌性如选择短尾则使自己的儿子适应度降低。因此二者都不能脱离这个稳定点。如果自然选择不足以制止性选择的非适应进化的话，这种“逃离”效应会使种群缩小乃至导致物种灭绝。按照 Lande 的分析，“逃离”倾向的大小取决于雄性性状与雌性对该性状选择性之间的遗传相关 (*Genetic covariance*)，也就是说尾巴较长的雄性个体也携带着倾向于选择长尾的基因，反之亦然。

剩下的问题是上述的过程是怎样开始的，即多数雌性选择长尾的初始状态怎样才能从无选择状态产生出来。费舍认为性选择开始产生时一定是与自然选择方向一致的，由自然选择的进化压力推动了性选择的产生。这个观点是正确的，但还不够具体，例如我们怎样才能理解羽毛的鲜艳色彩，宛转的歌喉，长而无用的尾巴对鸟类是有利的性状呢？我想这可以分两个阶段来说，第一个阶段是由于鸟类的飞行能力使其活动范围加大到三维空间，从而使视觉、听觉这些远距离信息手段的重要性超过了嗅觉。色彩

和歌喉起初作为一种单纯的求偶信息被利用。第二个阶段可以用汉米尔顿的寄生学说来说明。由于寄生生物使其宿主缺乏剩余能量来实现色彩和歌喉这种夸张性状，因而这些性状可以作为雄性携带有对抗寄生生物基因的证明，雌性选择这种性状将使下一代获得抵抗寄生生物的能力，提高适应度，因此对于夸张性状的选择得以超过单纯传递信息所需要的限度，进一步进化，实现“逃离”所须的初始条件。

纵观上述自始至终的全过程，我们可以对于性选择所引起的逃离效应有一个更具全局性的理解。从地形图模型来看，在自然选择的压力之下小球从坡上向坑底前进，如果在前进过程中产生了性选择的逃离效应，那么在到达坑底之后，小球会由于逃离机制继续向前运动，脱离最佳点，呈上坡之势。这个过程看起来象是小球有了惯性，可以称之为逃离机制所引起的“进化惯性”。为了更广义的讨论这个问题，我们把象性选择这样一种由系统（种群）内部因素决定的选择称为“内选择”，而象自然选择这种由系统的环境所决定的选择称为“外选择”。内选择可以决定一个内地形图，外选择可以定出外地形图，两个地形图合成在一起构成总地形图。那么所谓的逃离效应就是由于内地形图梯度的影响，使总地形图的坑底与外地形图的坑底发生偏离，甚至使总地形图成单方向倾斜，小球滚出界外，系统崩溃，导致物种灭绝。这可以说是由于性选择的发明导致地形图发生变化的一个例子。此时从外地形图上看，小球象是停在坡上不动，甚至向坡上滚去。

逃离效应到底使现存的动物在外地形图上多大程度偏离最佳点，有多少物种因逃离效应而灭绝，对于这些问题我们还缺乏充分的研究。我们更感兴趣的问题是，在文化的进化之中，逃离现象似乎更普遍地存在着，可以对我们常常听说的“异化”现象作出一种解释。文化中的价值观、道德、审美意识经常处于超越于利害和不受审察的位置，这些因素所造成的内选择也完全可能使

社会偏离最佳点，从维护社会存在的因素异化成给全社会造成损失的因素，甚至导致某些文明的灭绝。缠足和高跟鞋只是逃离效应的两个小例子而已。另一方面，从外地形图上看，逃离机制可以使小球向坡上滚，这意味着可以使进化过程翻越山口，产生了一个从较浅的坑转移到另一个更深的坑去的可能性，也就是脱离局部极小点的方法。形而上学是极强的内选择因素，现代科学产生于形而上学，从神学的婢女进化为反形而上学的独立力量，这个过程也是一个翻越山口的过程。在翻越山口之前的爬坡阶段，科学须要形而上学的逃离机制作为推动力，一旦翻过山口，即科学自身的有效性在外部表现出来之后，科学可以从实用角度产生自身的价值观，即创造了新的地形图。此时科学不再是证明上帝之伟大所使用的工具，形而上学的内选择作用减退，科学在自身的发明所造成的外地形图中作下坡运动。下坡运动的方向是明确的，由此可以产生新的内选择价值观，对总地形图进行修改，为将来的爬坡作了准备。当小球到达新的坑底时，科学的发展呈现暂时的停滞，此时各种内选择因素开始起作用，沿各种不同的方向作爬坡的尝试。此时各种形而上学性的东西受到重视。人工智能中的符号主义、逻辑主义、连接主义、行为主义、神经达尔文主义等都是爬坡的内选择价值观。在爬坡过程中如果发明了新的地形图，从而翻越山口找到下坡的方向，科学又进入一个新的发展期。需要注意的是，另一个更深的坑并不是原本就存在的，而是发明的产物。因此用“退火算法”的初始高温状态或遗传算法都是找不到这些坑的。这是地形图模型最容易造成误解的地方。以上的分析如果改用软硬结构模型来作，就不会引起这种误解。但是地形图模型有其形象化的优点，易于形成直观的认识。

3.7 广义进化范式

本书主要的目标是研究人脑的认知与思维过程，对于认知过程给出机械论的解释，也就是研究认知和思维的机制，使其有可能在机器上实现。目前的机器与人的最大区别在于没有创造性。因此如何实现创造（发明）即寻求可以在机器上实现的创造机制也就成为研究认知过程的关键问题。自然界中已经存在的最富有创造性的过程是进化过程，而且进化论又是一个机械论的理论，进化不需要依赖任何外在意识或目的论的帮助。达尔文所提出的进化机制是变异与选择。阿希贝的超稳定系统就是这个机制的机器实现方案之一。超稳定系统中的随机信号发生器对应着变异，进入稳定状态时关闭随机信号就是选择，稳定则是选择所必要的价值观，与存在原则一致。但是用超稳定系统解释智能时遇到重大的困难，这就是指数爆炸。这个问题也暴露了达尔文进化论本身的不完全之处。指数爆炸问题至今还是神创论者攻击进化论的武器。指数爆炸是指随问题规模或难度的增长其解决所需的时间呈指数增长，这意味着用变异与选择机制不可能一次性的作出大的或复杂的创造。但是把指数爆炸倒过来看，随问题规模和难度缩小其解决所需的时间呈指数函数减小，也就是说只要问题的规模和难度缩小，一定可以在任意短的时间解决。这意味着变异与选择机制可以进行小规模、简单的创造，阿希贝实际制作的系统就是一个小系统。余下的问题就是寻找把小创造积累成大创造的机制，这是我们建立广义进化论的动机之一。软硬结构模型就是这样的积累机制，硬结构是软结构创造性建构的积累结果，同时也是软结构走向更高级创造的支撑。

在本章中我们应用广义进化论的思想在进化生物学中提出了关于硬信息量的可验证假说，用以对进化重演律、进步性的进化、

有性生殖的二倍成本问题等进化生物学中悬而未决的问题给出更深层次的解释。通过这些理论，一方面可以展现广义进化论在生物学领域中的理论力度和深度，另一方面也作为用广义进化论观点建立理论的范例，使读者通过实例领悟广义进化论的精神实质，掌握分析问题的方法。此外在本章中也介绍了一些达尔文进化论的现代研究成果。达尔文进化论是广义进化论的前身和必要组成部分，这些研究成果对于广义进化论也同样是重要的，为我们提供了用进化论方法分析问题所需要的模型和方法。在介绍这些模型时我们也注意用广义进化论的大时间尺度眼光来看问题，对于“逃离效应”给出整体连续的而不是一次性的图景，挖掘出“逃离效应”更深层的意义。用相同的观点，对于囚犯困境也可以有一个更具整体性的认识，把阿克赛洛德的解与哈丁的解连接起来，就可以构成一个整体的解，而哈丁解是要靠内选择维持的，也存在着“逃离”的可能性。这些模型的一个重要意义是给出价值观生成的机制，使价值观得到具有客观性的依据，以解决丑小鸭定理所造成的主观化困难。进化离不开变异与选择，而选择离不开评价，如果没有一个具有客观性的价值理论体系，进化论向生物学领域以外进行推广是缺乏理论基础的。在生物学领域之内，达尔文的适应度可以作为评价的基础，出了这个领域问题就变得复杂起来了。这也就是我们为什么有必要在第二章建立以价值为核心的哲学基础的原因。存在原则就是“优胜劣败、适者生存”在哲学层次的对应概念。囚犯困境与逃离效应给出了由基本价值观生成次级价值观的过程和机制，也说明次级价值观对于说明大时间尺度进化过程的意义，以及次级价值观在内选择中的爬坡效果和逃离的危险性。从逃离效应的全过程来看，逃离的危险性产生于这样一个现象：价值观是一个向量而地形图却不是个超平面，而是由多个非线性因素综合而成的超曲面。因此对于在某一点上与地形图下降梯度一致的价值观，在超过最低点之后必定成为负价

值，如果这个负价值具有自我保持的力量，就会造成逃离的危险性。例如从达尔文的适应度出发，人生育得越多越好，但是地球的有限资源不可能承受指数爆炸的人口，生态环境的破坏可能导致人类的灭绝，因此限制人口成为一个重要的价值观。有一种极端的观点认为，地球是一个生物体，人是这个生物体上的癌症，为了拯救地球，应该让人类灭绝，使地球永保持其绿色。这个观点是明显荒谬的，如果人类灭绝了，又有谁会认为绿色的地球比红色的火星更好呢？绿色本身又有什么意义呢？好在这个理论本身是违反存在原则的，不可能被多数人所实践，但是这个理论在人口过剩的情况下，依然有其价值，会被一些人接受。

广义进化范式与欧几里德范式相比，一个重要区别是前者是基于价值的而后者是基于真伪值的。因此我们所重视的不是某个定理真伪的证明，而是一组非线性评价函数的组合优化，以及价值的基础理论，价值在进化中的生成过程与作用，多重价值观的权衡等等。值得注意的是，进化论本身作为一种理论是不具备可证伪性的，这一点与神创论相同，但是进化的思想为我们带来的科学成果比神创论多得多，这就是进化论优于神创论的理由，而这个理由也是自我表述的，符合进化论的只谈优劣、不论真伪的原则。

广义进化范式的第二个特点是强调非决定论的一面。按欧几里德范式建立起来的牛顿力学是强调决定论的，利用牛顿力学我们可以精确地预测下一次日蚀在何时何地发生。但是由于变异所固有的随机性，使得进化成为原则上不可预测的过程。我们不可能根据进化论来预测下一次将在何时何地产生什么样的新物种，哪怕是预测流感病毒这种最简单的生物也作不到。从预测原则来看，非决定论是令人失望的，但是正是这种不可预测的性质使得广义进化范式可以用来建立含有创造或发明的理论，而这种理论用强调决定论的范式是不可能建立的。这是因为创造或发明的含

意就是不可预测的，如果我们能够精确地预测明年的某一个发明，那么我们就已经作出了这个发明，它就成为了现在的发明，而不再是明年的发明。对发明的预测本身就是发明，因此发明和创造是不可预测的。

广义进化范式的第三个特点是重视演化的动态过程，广泛使用“似目的论”的观点。欧几里德范式是追求不变性的范式，比较注重静态结论，因此所提的问题通常集中于“是什么”“怎样”和“为什么”。而进化论中所提的主要问题是“从何而来”、“有什么好处”或者“为什么能够存在”之类。在欧氏几何之中，定理绝不能用其有用性作为证明的依据。在牛顿力学中，“月亮的存在对地球有什么好处”是一个没有意义的问题。在欧几里德范式之中，目的论与机械论是完全对立的。但是在进化论之中，物种从何而来的动态过程成为重要的问题，达尔文提出进化论的书名就叫《物种的起源》。以进化论的眼光来观察事物，促使我们对生物的各种性状提出一个共同的问题：这种性状在进化过程中为什么能被保存下来？这个问题更简化的表现是，这种性状对生物有什么好处？这种简化提法在形式上是非常接近于目的论的，但它又有变异与选择这种机械论的机制作为基础，因此我们称之为“似目的论”。从“似目的论”的观点出发，我们在研究认知与思维的各种现象时，不仅仅注意这些现象本身的机制，更重要的是研究这些机制在进化和学习过程中是怎样产生的。思维能力提高的过程比思维过程本身更重要，因此我们把学习和创造这样的动态过程作为关心的重点。对于象情绪和意识这样的高级中枢神经活动，人们往往容易产生形而上学的看法。注意到这些现象是进化的产物，追寻其生物学目的，可以使我们避免形而上学，认识其生物学功能，以此为线索有助于找出实现功能所需的结构和机制。

广义进化范式的第四个特点是其大时间尺度的观点，这也是广义进化论超越达尔文进化论之处。广义进化范式不仅重视一次

性的小创造,更重视由多个小创造产生大创造的积累过程。发明、调整、再发明的循环,软硬结构模型,地形图与价值观的相互生成等都是对这种积累过程从不同角度的探讨,其共同特征是都有某种循环的结构,通过循环打开通向开放和无限的道路,多个循环构成进化的链条整体。按欧几里德范式所建立的理论大多呈现有始有终的直线结构,因此所描述的现象也是一次性的。协同学、耗散结构、突变论等都属于这种一次性的理论。一次性的理论主要用于单一问题的求解,其条件和问题空间是固定的,不考虑解本身对问题空间的影响,也不考虑解创造新的问题空间的可能性。通常我们考虑变异和选择时,只考虑选择对变异的约束作用,不考虑变异产生新的选择评价准则的可能性。在这样的模型中,进化地形图是固定的,遗传算法虽然引进了进化机制,但是没有超越固定地形图的框架,其效果也就不显得比其他优化算法有明显的提高。这是一种注意外因的一次性模型,系统的问题是由外部给定的,求解之后就停下来等待下一个问题。广义进化论给出了不断进化和进步的范式。注意分析问题的解和求解过程产生新问题的机制。问题的产生与求解形成环状结构,从而把侧重点放在系统的内因上。

广义进化范式是以价值评价为基础的,它本身也代表了一组评价准则:注重评价及价值的权衡,注重动态的发展和演化,强调非决定论,注重不可预测的创造性过程,注重环状的理论结构,注重大时间尺度的整体考虑,注重变异与选择的相互作用,注重问题的生成过程,注重某种性状的起源和好处,强调似目的论的方法论,注重系统演化的内因,重视软硬结构模型,重视地形图的变化内因,重视发明与调整的周期性积累,注重多样化的长远和全局价值,等等。总而言之,重视其他的科学范式所经常忽视的方面,以解决由于这种忽视所带来的理论困难,找到分析和综合自我表述系统的方法。

第四章 认知与思维

4.1 认知的阶层性

人脑是一个复杂而巨大的系统，认知与思维过程也是极为复杂的过程。为了理解一个复杂的大系统，我们通常的作法是，先把一个复杂的大系统分解成几个复杂程度较低的子系统，研究子系统的功能和各子系统之间的关系，从而达到理解大系统的目的。所谓总体大于部分之和，就是指只由子系统的功能不能完全说明大系统的功能，还需要加上子系统之间的关系，这个“关系”就是“大于”所多出来的部分。零件的堆积不是机器，零件按一定位置装配好才能产生机器的功能，这一定位置就是零件之间的“关系”。这就是科学分析的方法，或称为还原论的方法。如果通过分析得到的子系统都容易理解，它们之间的关系也是明确而可以理解的，那么这个分析就算大功告成。如果得到的子系统仍然不容易理解，我们可以把该子系统看成一个大系统，进一步分成更小的子系统，重复上面的分析过程，直到完全理解为止。如果在分析过程中出现一个既不容易理解，又不能进一步分析下去的子系统，那么我们可以说这是一个失败的分析。把系统分为子系统的方法是很自由的，但分析也受到明晰性的约束，如果分析所得的子系统或它们之间的关系难以理解，分析也不能说是成功的。对于一个极其复杂的系统来说，如果我们一下子把它分析成大量的基本子系统，那么由于子系统的数量太大，子系统之间关系过于复杂而难以理解，这种分析是难以成功的。例如把大脑看成是一百亿个神经元组成的系统对于理解大脑的帮助不大，把人看成

细胞的组合或把社会直接分析到个人都导致关系不清的结果，对于系统的理解和掌握是不利的。因此对于极其复杂的系统来说，先分解为少数的子系统，再分为孙系统、重孙系统的作法，更容易实现分析的明晰性。分析过程中的这种“辈份”关系就构成了系统的阶层性，这种阶层性靠指数爆炸的帮助，使我们有可能用少数的共同原理和机制去掌握极其复杂的大系统。这就是分析复杂大系统时阶层分析的必然性，可以称之为逐层的还原论。例如我们可以把人体分为消化系统、循环系统、神经系统、呼吸系统、生殖系统等各个子系统，各系统再分为器官，器官之下又有组织，组织再分为细胞。这样，人体就分为系统、器官、组织、细胞四个层次，比从细胞直接理解人体要容易理解得多。各层次有各层次的原理和机制，上层的原理可以由下层子系统的功能和子系统之间的关系等下层原理来解释。相邻的上下层之间，下层对上层的解释能力很强，但是对于不相邻的层，要想用下层原理来解释上层通常是很困难的，否则中间夹着的阶层就成为多余的层了。

对于象人脑的认知与思维过程这样的极为复杂的系统中的复杂功能，要想得到全局性的说明当然必须进行分层的解释，层次分得不够就会造成理论上的困难。我们把与认知有关的各种概念粗略地分为如图 4.1 那样的 7 个层次。7 个层次又可以归纳为 4 层。

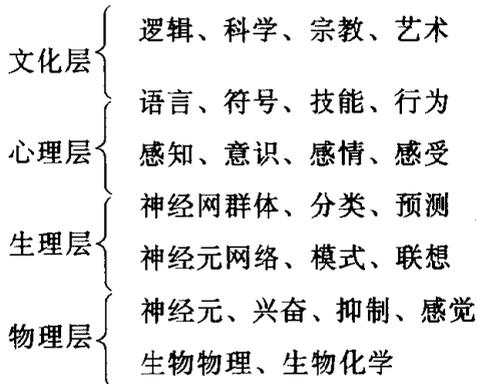


图 4.1

分层的基本原则是上层概念可以由其相邻的下层概念加以解释。由于有些概念涉及不止一层，各层次之间也有相互作用和影响，所以不应把这种分层看得过于绝对。这种分层只是给出大致的相对位置，说明这些概念之间的上下级关系。进一步的说明，有可能需要更细的分层，这里只是一个全局性的初步分析，以表现关于认知的大致轮廓。有些重要的概念，例如学习、记忆等没有列在某一层次之中，这是由于这些概念含意很宽，几乎涉及所有层次，在不同层次中有不同的机制，只能在各层次中分别说明。以下我们用计算机和电视机作为参照，说明各层次的大致内容和相对位置。

按自下而上的顺序，**第一层**是属于物理层的生物物理、生物化学层次，这一层对应着与计算机和电视机有关的说明晶体管工作原理的半导体物理学。这一层次已经有不少人作了大量的工作，例如研究神经冲动传导的电化学机制，研究突触权值变化的物理和化学机理，研究神经细胞兴奋与抑制的生物物理和生物化学机制等等。在这一阶层上与学习和记忆有关的是突触权值（与其他细胞的连接强度）变化的机制。这个层次由于与心理层次隔了几层，因此直接用物理或化学原理去说明认知是极为困难的，就象想用固体物理学来说明计算机和电视机的原理一样。有一些化学家或生物化学家试图从化学的角度来研究认知，从多年以前的“恐暗素”实验，到后来的一些关于认知研究的从生物化学立场的探讨，例如 Michael Conrad (1986) 所提出的用酶与蛋白质之间的特定结合来解释模式识别的分子信息处理模型。虽然蛋白质与酶的种类有可能多到足以对我们的概念全部进行编码，但是这种分子学说由于其非阶层性，对于说明更复杂的心理活动是很困难的。

第二层涉及生理层和物理层，中心内容是研究单个神经元的特性、神经元与其他神经元之间的信息传递以及神经元内部的信

息处理。其研究领域属于神经生理学和人工神经网络的基础部分。参照计算机和电视机，这一层相当于单个晶体管的工作原理。从神经生理学角度对神经元进行研究，往往得到一些较为逼真的数学模型，这些模型通常比较复杂，以至于可以说一个神经元的功能就可以相当于一个微处理芯片。但是从分层解释的角度来看，过于复杂的模型用于对上层（神经网络）的说明是不适当的。因此我们比较倾向于采用人工神经网络研究中所用的尽可能简化的模型，例如图 4.2 所

示的模型。其中的 x_1, x_2, \dots, x_n 代表神经元各树突的输入信号， w_1, w_2, \dots, w_n 代表各个树突的权值，通过神经元的兴奋，在轴突上产生的输出信号 y 可以用下式表达：

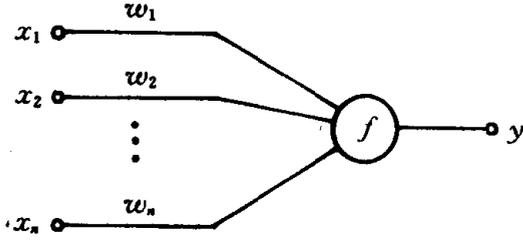


图 4.2

$$y = f\left(\sum_{i=1}^n x_i w_i\right)$$

 y 是各输入与对应权值的乘积之和的函数，函数 f 可以是阈值函数或是 Sigmoid 函数。当然，神经元的数学模型还有很多，我们选择这一个作为例子只是因为它比较简单而又能够说明很多问题，在第二章里我们曾经用这个模型说明过权衡利害的“中庸算法”。在这个层次上学习和记忆表现为各权值在神经元工作过程中的变化，关于神经元学习的算法很多，我们只举一个经典的 Hebb 学习算法为例，这个算法是 D. O. Hebb 在 40 年代提出的，其他算法大多是这种算法的改良。用图 4.2 中的模型和符号，Hebb 学习算法可以表现为：

$$w_i = \sum_{t=0}^N y(t)x_i(t)$$

Hebb 学习准则的基本意思是，如果两个神经元同时兴奋，则连接这二者的突触的权值就应加大。上面的式子是由这个原则积累得到的。式中的 t 代表次数。Hebb 学习准则及其改良版足以说明其上一层的神经网络的学习算法。

第三层是属于生理层的神经网络层次。位于这个层次上的研究领域有神经生理学和人工神经网络两方面的研究，这两个方面密切相关，互相参照，相互补充，相互验证。这个层次可以对应于计算机或电视机中的基本电路层次，例如门电路、触发器、放大器、振荡器等等。仍然是出于说明的简单性要求，在这一层我们以人工神经网络作为主要内容，适当参考神经生理学的成果。在这一层次上学习与记忆主要表现在神经网络的一些基本功能，例如自组织功能、映射与函数逼近功能、联想存储器功能等等。

第四层是神经网络群体层，关于这一层以及再向上的层次研究得还很不充分。神经网络群体是一个比较宽泛的概念，群体本身可以具有复杂的层次和模块结构。这一层相当于计算机中的存储器、运算器、控制器、接口等各功能部件，或电视机的高中频线路、视频放大、音频放大、彩色解码、扫描线路，显象管等功能部件。由于群体所具有的多层结构，这一层所说的群体的结构与更上一些层的功能有关，而这些功能有很多是目前还没有完全弄清楚的，因而我们暂时不在结构上作进一步分层，只是用一个宽泛的群体概念来代替更具体的结构分析，而在以上层次中主要讨论功能的分层，为实现这些功能所需的结构分析准备前提条件。这一层次涉及心理层与生理层。在功能方面，这一层讨论心理层中基本层次的功能，例如分类与模式识别，动作模式生成、局部预测、智能模型（Mental Model）等等。学习和记忆则体现在这些功能的生成过程之中。

第五层可以称之为意识层,这一层处于心理层的中心位置,所研究的主要问题与心理学的研究层次基本一致。这个层次对应于计算机或电视机的整体结构。虽然在意识层中主要从功能入手进行分析,但其分析的结果将对于构筑高度智能系统的结构和理解大脑的组织结构具有指导意义。也就是说意识层对于智能系统的总体结构起着决定性的作用。意识层所研究的主要问题是意识、情绪、知觉、行为、思考等高级神经活动,并力图用低一层次的原理对这些活动加以解释,从而通过分析得到构筑具有这些功能所需结构的线索。学习和记忆在这一层次中表现为意识对于学习过程中指数爆炸的回避作用,情绪评价对于学习和记忆的指导作用、知觉对于复杂学习过程的必要性,行为的学习过程、思考方式在学习过程中的变化以及情绪评价通过学习而变化的过程等等。

第六层是语言层,这一层次涉及心理层和文化层,与认知系统的硬件结构有关,但影响不象意识层那样大。语言层对应着计算机的操作系统和电视机的制式,这些都与硬件有一定的关联,但又对硬件有一定的独立性。这一层主要的研究对象是语言、符号表现、技能的发达等等。符号表现是语言概念的推广,包括姿式、图形、表情等各种表现手段,以及对这些表现的理解能力,语言只是符号表现的一个特例。技能是指经过复杂组织的动作而产生的熟练行为。符号表现能力或语言的使用也是一种技能。掌握语言的能力与硬件有关,依赖于大脑皮层的语言中枢,以及操作发声器官的运动中枢,但是说哪种语言却又取决于文化环境。因此语言能力是生物遗传的,而具体的语言结构则产生于文化进化。包括符号表现在内的其他技能也是一样的,一方面依赖于认知系统或脑的硬件结构,一方面又取决于文化环境。如同操作系统是计算机软件与硬件之间的接口一样,语言层是心理层与文化层之间的接口。在语言层中关于学习和记忆体现于心理层中一些原理的具体应用和复杂的组合,以及文化层与心理层在学习方面的关联。

第七层可以称为逻辑层，完全属于文化层次。这一层相当于计算机的应用软件或电视系统中的电视节目。文化层中的问题与硬件系统的关系较弱，原则上计算机的软件可以在不同结构的计算机之间移植，电影或舞台剧也可以在电视系统中播出。属于逻辑层的内容包括逻辑、科学、宗教、艺术等所有文化现象。学习和记忆在这一层里表现为文化的进化过程、个体认知与文化环境的相互作用、高水平的发明创造，等等。把这一层称为逻辑层是有其特殊用意的，它强调逻辑是文化进化的产物，是认知的上层建筑，而不是其基础。语言层的名称也具有类似的意义，即认知是语言和符号的基础，对认知的心理层的理解有助于说明语言、符号和逻辑，但反过来用语言、符号和逻辑推理来说明认知和思维过程则是极为困难的。虽然乔姆斯基早就指出语言是基于人的天生的能力，但直至今日我们对于这种天生的能力还几乎是一无所知，没有找到可以支持所有语言的人类统一语法。其原因就是对更深层次的认知与思维过程不够清楚。

有了这样一个粗略的分层，我们在讨论与认知有关的具体课题时，就可以知道这个课题在整个认知问题中的位置，这个课题与其他课题之间的关联。例如为了研究这个课题需要先解决哪些其他课题作为必要的预备知识，这一课题的解决又应为哪些更高层的其他课题准备条件，因而这一课题中哪些研究方向和结论是重要的。通过理解某个课题在认知科学中的位置，才能定出关于这个课题的有效的研究方针。通过分层，使我们对与认知有关的各个研究领域在整体理论中的位置及重要性，以及它们之间的相互关系和研究的先后顺序有所认识。从认知科学的现状来看，我们认为当前最困难因而最引起我们兴趣的问题是心理层的问题，这是由于心理层以下的生理层和物理层已经形成了发达的研究领域，研究的方法比较完备，成果也很多，而心理层则基本上还处于盲人摸象的局面，只有一些零散的猜测，未能形成体系，因而

使得心理层以上的层次研究起来比较困难，缺乏必要的基础。

按照控制论、信息论和系统论的基本观点，我们对于心理层的认识偏重于其信息模型，或者说是数学模型。因此物理层次中的物理、化学原理对我们来说重要性就不太大。一旦通过物理或化学原理的分析得到信息模型，物理和化学原理本身就可以舍弃了。例如某种信息是通过物理的方法还是化学方法传递，在信息模型中只表现为传递的速度和范围，物理和化学现象本身并不重要，是可以互换的。例如从信息模型的角度看，只要能作出逻辑的与非门，就可以组成计算机，至于与非门利用什么样的物理原理来实现，对于计算机的原理是没有什么影响的。例如可以用凸轮等机械装置，也可以用继电器、气体或液体的阀门、射流元件、电子管、晶体管、磁放大器、生物神经元、光器件、超导器件等等，物理层的变化只体现于价格、体积、速度等因素，对于计算机的信息模型是没有影响的。物理层与心理层之间隔着生理层，因此直接影响很小。某种信息模型怎样在物理上实现，是一个技术与经济问题，不是信息理论上的问题，而一个信息模型是否可能在物理上实现，才是重要的信息理论问题。由于物理层与心理层没有直接关系，而且这方面的研究已经比较成熟，所以本书不准备讨论物理层次的问题。

4.2 神经网络

生理层的研究主要由神经生理学，心理学和人工神经网络几个领域组成。由于我们对生理层的要求倾向于对更高层次的说明，因而比较关心尽可能简化的模型，所以较偏重于人工神经网络的一些研究成果。人工神经网络（以下简称“神经网”）的研究发端于40年代，1943年心理学家 McCulloch 和数学家 Pitts 合作提出了形式神经元的数学模型。现串行机的奠基人冯·诺依曼 (Von

Neumann) 在 50 年代就已注意到计算机与人脑结构的差异, 对类似于神经网的分布系统做了许多研究。50 年代末, Rosenblatt 设计了感知器 (*Perceptron*), 试图用神经网络模拟动物和人脑的感知和学习能力, 这是从工程上实现神经网的早期例子。1962 年, 人工智能创始人之一的明斯基 (M. Minsky) 和 *Papert* 以感知器为名出版了一本书, 从数学上深入分析了感知器的原理, 指出了它的局限性, 加之当时串行计算机正处于全盛发展期, 早期人工智能也显示了“成功”的兆头, 使得神经网络的研究一度处于低潮, 当然在这个低潮时期, 也有一些人作出了出色的研究。从 80 年代开始, 经典的符号主义人工智能的局限性逐步明朗化, 而神经网络的研究又出现了一些重要的突破, 例如 Hopfield 网络声称可以解决著名的“旅行商最优路径问题” (*Travelling Salesman Problem, TSP*), 为研究者展现出征服指数爆炸的诱人前景。1986 年 Rumelhart 和 McClelland 提出了多层神经网络的反向传播 (*Back propagation, BP*) 学习算法, 打破了明斯基的结论所造成的悲观气氛, 使得大批的研究者涌入了神经网络的研究领域, 从而迎来了神经网络研究的全盛时期。时至今日, 我们可以说, 神经网络的研究已经进入了成熟的阶段。这里所说的“成熟”并不意味着已有的研究成果十分令人满意, 我们可以用神经网络为所欲为, 而是意味着我们对于神经网络可以干什么和不可以干什么有了比较清醒的认识, 对于神经网络理论上的局限性已经有所理解。从全局的和普遍的角度看, 在神经网的领域里, 我们仍然匍匐于指数爆炸的雷区之中。明斯基的悲观结论基本上是正确的, 在神经网络中我们所关心的几乎所有问题在普遍意义上都是指数爆炸的 (在专业术语中称为 *NP* 完全问题及 *NP* 困难问题)。TSP 问题的解法没有普遍意义, BP 算法对于大规模的复杂问题同样无能为力。这种悲观的结论仅仅说明, 只用神经网的原理去实现高度的智能是不可行的, 不应该把人脑看成一个均匀的巨大的神经网络。反之, 神经网络作为

认知系统的基本原件，以支持更上层的理论则是可行的，此时我们只需要对神经网络以较低的要求，即解决小规模简单问题，而把由小智能积累成大智能的任务留给更高层次的理论去解决，那么我们就用不着和指数爆炸正面冲突了。从这一点出发，我们在神经网络领域中所关心的问题就会有一些特色，与一般的神经网络研究的重点有所不同。

我们所关心的一个问题是，在神经网络中处理的基本信息形式是什么。对此有两个可选择的方案，一个是模式，一个是符号。人脑的神经网络中所处理的信息到底是模式还是符号，这种本体论的提问方法是没有多大意义的。对于神经网络中的同一组信息，我们都是既可以把它看成符号，也可以看成模式的，问题仅在于怎样看对于理论的展开更为有利。由于这种看法的不同，就出现了符号主义和连接主义、逻辑主义和格式塔学派的区别。例如，对于一个神经元，我们既可以把它看成完成某种逻辑运算的门电路，也可以看成对于某一类特定模式起反应的分类器。对于一个字母 A，既可以看成是一个符号，也可以看成一个特殊的图形，加上一定的读音，图形和声音都是模式。字母“A”在计算机中用 ASCII 码表现为十六进制的 41，可以把它看成一个数字（符号），即十进制的 65，也可以看成一个二进制的 0 和 1 的排列，即 01000001，是一个模式。从认知的角度出发，我们在神经元和神经网络这个层次上倾向于用模式作为基本信息形式。这种选择出于以下几点理由：第一，我们从外界直接获得的信息是模式信息，如果认为思考是一个符号处理过程，二者之间的变换，特别是变换过程的学习将会遇到很多困难，这些困难实际上是由符号处理这个多余的假定造成的，我们见到一个字而读出它的发音只需要完成从图形模式到口腔及声带肌肉运动模式的变换，中间并不一定需要经过模式与符号的变换与反变换过程。第二是模式之间存在着自然的类似关系，例如可以在模式空间中定义欧几里德距离或是模式向量夹

角余弦之类的类似关系，从而定出所有模式之间的“自然的”相似程度。虽然这种“自然的”相似程度未必符合认知主体的主观要求，但可以以此为基础逐渐修改到合乎要求。但是对于符号来说，原则上不存在“自然的”类似关系，符号之间的一切关系要依赖人为的定义，因此以符号为基础的系统，通常难以具备自组织和自学习的能力。当然我们选择模式作为信息的基本形式，并不仅仅出于以上的两个浅层理由。更深层的背景是，我们认为人和其他生物的生存目标并不是追求合乎逻辑的真理，而是追求生存本身；思维的基本任务不是认识客观事物的本来面目，而是找到最佳的活法；思考的基本方式不是逻辑演绎，而是利害的权衡。有人认为基于符号和逻辑运算的计算机之所以显得比人脑僵化和死板，是由于人所常用的概念具有模糊性，例如年青和年老的区别，热水和温水的区别，都没有确定的分界，从而主张采用模糊(Fuzzy)逻辑以使计算机接近人脑。但是人脑与符号系统的最重要的区别并不在于概念的边缘是否清晰，而在于概念的相对性和主观性，例如人们常随着自身的年龄增长而把划分年青与年老的界限提高，人的皮肤的皱纹多少也会使人看起来显老的程度与实际年龄有很大差别，象这种多因素的相对概念用模糊逻辑处理并无太大的优越性，因为模糊逻辑与普通的二值逻辑一样属于绝对的体系。权衡是多因素的相对比较，而模式是表达多因素信息的有效形式。

从认知的分层理解的立场来看，上述“连接主义”倾向只是我们在生理层和心理层所持的态度，对于其他层次可以视需要而采用不同的观点。例如在生理层之下的物理层中，神经元及神经网络在原则上都可以用理想的图灵机或现实的冯·诺依曼型计算机在任一精度上进行完全的模拟。因此，“图灵机可以模拟人脑”这个命题可以接受。在这个意义上，我们有理由把人脑看成一个“物理符号系统”。在心理层之上的文化层中，例如当我们讨论语

言与逻辑的关系时,把语言看成一个符号系统当然是很方便的。经典人工智能和“符号主义者”们的失误并不在于对符号的重视,而在于对符号所组成的其他更复杂的信息形式的无视。由于看不到生理层与心理层的存在,把文化层直接还原到物理层,导致了不象智能的人工智能。举一个简单的例子就可以说明文化层与物理层的区别:计算器可以进行加减乘除的计算,儿童在学过算术之后也可以作算术运算,谁也不会否认作算术题对于人来说是一种智能活动,但是计算器的算术运算与人的算术运算显然具有不同的原理。谁也不会认为儿童在学过算术之后在大脑的某个部位组织出一个与计算器相同的神经电路。人学习算术是一个复杂得多的过程,首先是一个文化过程,学习算术需要文化环境,起码的文化交流能力(广义的语言),还要有心理因素:必要的基础知识,必要的理解能力,学习的愿望,记忆的能力,以及掌握技能(算法)的能力等等,所有这些,都是计算器所不具备的。从这种层次的区别,我们可以知道人所进行的逻辑思维与计算机的“逻辑思维”在原理上是不同的,人的逻辑思维是一种文化现象,依靠着心理层的形象思维(模式联想)的支持,所谓逻辑思维通常是指运用语言组成的智能模型所进行的思维过程,这种过程与计算机中的逻辑运算共同之处不多,可以说,人不能够进行计算机那种“纯粹的逻辑思维”。因此,我们不能认为造出计算机就意味着掌握了人脑的逻辑思维的规则,正如造出计算器不意味着知道人作算术的过程一样。由此可知,所谓人的逻辑思维是建立在形象思维基础之上的一种后天技能,把人的思维看成是逻辑思维与形象思维两种同等地位的思维方式的组合,甚至根据某些实验断定左脑和右脑分别进行形象思维和逻辑思维,都是过于简单化的想法。从认知的阶层理论出发,我们就很容易理解诺尔曼所说的“我们确实是个物理符号系统,但不仅仅是个物理符号系统”这句话的具体意义。也就是说,认知系统最终一定可以还原成物理符

号系统，但是这种还原必然是多层次的，需要通过很多不同层次的概念才能逐步完成。要想理解人的认知过程，只有符号处理的机制是不够的，还要加上其他各层次的机制。

从认知的阶层性理论出发，我们可以探讨在生理层次之中，为了支持上层的功能，神经网络应具有哪些最低限度的基本功能。或者说，在已经经过详尽研究的众多的神经网络模型和功能之中，我们选择哪一些作为支持上层机制的基础。从原则上来说，为作为基础而选择的模型或功能应该是尽可能简单而通用的。用细胞来作比喻的话，研究单细胞生物的细胞与研究复杂生物组织中的细胞所关心的重点不同，象草履虫这样的单细胞生物，我们需要知道它的摄食、消化、排泄、运动方式等一系列功能，但是对于人的肌肉组织中的肌肉细胞，我们通常只关心它接受神经信号收缩的功能，这样已足以说明肌肉的运动。这就如同在机器的总装配图中，我们不必去表现零件的细节，只着重表现零件与整机工作原理有关的功能和各零件之间的关系。按照这样的简化原则，对于神经网络我们只讨论以下的四个功能：模式联想、函数逼近、上下通讯、串并行变换。

在讨论模式联想的功能之前，我们先对模式的概念作一个简单的说明，模式通常是指具有一定的空间或时间分布的信息，当然也有一些模式并不一定在实际的时空之中，但也总可以在人为的时空之中给以表现。这个特点是与符号的最大区别，通常我们认为符号是不具有时间和空间分布的，如同几何学上的点。当不止一个的符号连成一串时，我们可以把这个符号串看成一个模式，因此符号与模式是密切相关的。模式的另一个特点与分类和识别有关，即模式往往带有某一类的时空分布信息的含意，这个问题比较复杂，放在模式识别的部分讨论。对于具有时间或空间分布的信息，往往可以用时间或空间坐标的函数来表现。由于神经元的离散性，加之任何认知主体的感觉信息输入都不可能也不需要

有无限的分辨率，对于函数表现的模式可以进行离散化而不失其普遍性，即在时空间中只取有限个坐标的表现。把有限个坐标与自然数对应，模式可以表现为函数值的有顺序的集合，可以称之为数组或有限维向量。当数组中的各元素取不同的值时就表现不同的向量，即不同的模式，所有向量的集合称为向量空间，或称为模式空间。对于一个神经元来说，各输入端的输入值可以组成一个模式（向量），各输入端的权值也可以组成一个模式。对于神经网络来说，各神经元的兴奋度分布也可以看成模式，各输出端的输出值也构成一个模式。

模式联想的功能可以用联想存储器来实现，联想存储器可在两个或两个以上的模式之间建立稳定的联系，当神经网络输入某个模式或与之相近的模式时，在输出端可以得到与之相对应的另一个模式，就完成了模式联想的功能。在神经网络中建立模式联想能力的过程称为学习过程，也就是联想存储的过程，即通过学习把模式联想关系存储到神经网络之中，此时神经网络相当于一个联想存储器。几乎所有类型的神经网络都可以用作联想存储器，例如感知器，BP网，Hopfield网等等，因此可以说建立模式联想是神经网络的一个基本功能。各种神经网络的学习算法有所不同，这里不作详细介绍，只考虑模式联想的共同特性。联想存储器与通常计算机中的存储器最大的不同之处在于，它不象计算机的存储器那样靠地址来寻找内容，而是靠内容来寻找内容，即所谓的内容寻址，从一个模式联想出另一个模式，而且允许输入模式含有一定的噪音或缺，而计算机的地址是不允许有丝毫差错的。

我们注意到一个数组可以切分成若干段，成为几个短的数组，若干个数组也可以连接起来构成一个长的数组，那么一个联想存储器的输入或输出模式也都可以分为若干段，赋予不同的意义。输入中的任一段模式的改变，都可能影响整个模式，使输出发生变化。这意味着可以由输入的某些部分控制整个联想的结果，这些

部分可以是场合 (*Context*) 或者情绪之类的模式信息。如果我们存储的模式联想关系中输出的模式与输入的模式相同, 就构成了自联想存储器, 此时输入一个有失真或残缺的模式可以得到正确而完整的模式, 从而可能用任何一小段模式的输入把整个模式的其他各段联想出来, 这意味着一种双方向或多方向的联想, 也就是把几个模式连接成一个大的模式存到自联想存储器中, 那么输入其中的任何一个或几个模式, 都可以把其他模式联想出来。

对于存储了若干对模式联想关系的神经网络来说, 输入若干个模式, 将会联想出若干个输出模式与输入模式一一对应。这种关系可以看成是一个函数关系, 也可以称之为映射或变换。从这个观点出发, 神经网络的学习或联想存储的过程可以看成是函数逼近或生成映射关系的过程。联想存储与函数逼近的机制相似, 但这两种看法体现的要求有所不同。联想存储往往要求存储的多个联想关系之间尽可能隔离而互不干扰, 即强调各采样点的孤立性, 或者说是侧重分化的; 而函数逼近则往往侧重于泛化, 力求用较少的采样点逼近函数, 而可以自动进行内插和外插, 使未知的输入模式也能得到适当的输出模式。如果所要逼近的函数比较简单, 例如线性变换, 学习是很容易的, 但是如果要是逼近的函数是复杂的非线性变换, 虽然从理论上多层神经网络可以实现任意复杂的非线性变换, 但其学习过程是指数爆炸的。

如果我们把神经网络看成是大规模的神经网络群体的一个部件, 神经网络之间的通讯就显得比较重要了。为了把这种复杂的通讯关系整理得较为清晰, 我们把神经网络的输入和输出各分为两大部分, 一部分是通向上层神经网络的, 另一部分通向下层神经网络。这样不管神经网络的内部结构和功能如何, 从外部通讯角度看, 它具有四类通道, 即向上层的感觉输出, 向下层的控制输出, 上层来的控制输入和下层来的感觉输入。对于这样的外部特性, 我们称之为四端神经网络。每一类通道都可以包括与不只一个其他神经网络的模

式信息通道，在多数情况下，这种通道是双方向的，即包括感觉与控制两个方面。对于神经网络之间的上下层比较，可以采用这样的原则：相对接近末梢的感觉和控制端而远离中枢神经的算下层，而远离末梢接近中枢的算作上层。就象火车车次的规定一样，从北京开出的火车为单号车次，是下行列车，而开回北京的是双号车次，算作上行。下行列车先到的车站可以定义为后到车站的上层车站。对于神经网络也可以定义上行的信息为感觉信息，下行的信息为控制信息。按这样的定义二环路上的车站都是同层的，但是可以在每次任务中分出不同的层次，例如有一辆车在从市内开往市外的下行途中在二环路上行驶了一段，那么在这个任务中可以认为先经过的站是后经过的站的上层站。这就象在计算机的操作系统中，控制显示器和键盘的程序与控制磁盘的程序同层，当键盘输入汉字需要参照词典而调用磁盘文件时，磁盘控制程序是下层，反之，当磁盘读写出现问题在显示器上显示错误信息，要求键盘输入处置方式时，显示器和键盘的控制程序是磁盘控制程序的下层。这是上下层关系依存于任务的相对性。

串并行变换不是神经网的典型功能，但是认知主体有时需要处理分布于时间轴上的模式，例如听觉、运动视觉和时间序列的动作等等，而神经网络通常处理的是同时的空间模式，因而时间模式与空间模式之间的串并行变换就成为认知系统所必须的功能。串并行变换最典型的模型是计算机中的移位寄存器，如图 4.3 所示。图中每一个方格代表一位的寄存器，可以存放一位数字，在时钟的驱动下，移位寄存器每收到一个时钟脉冲就从左边的串行输入端取一个数字放在 D_1 ，而把 D_1 中原有的数字存入 D_2 ，依此顺移， D_6 的数字从串行输出端送出。这样在并行输出口总是出现前六个时钟串行输入的串并行变换结果。如果把并行输出口改为并行输入口，在并行输入一组数字后，用 6 个时钟脉冲按同样原理可以在串行输出端得到并串行变换的结果。这样就完成了时间模

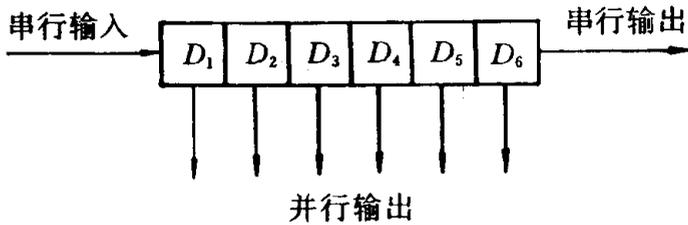


图 4.3

式和空间模式之间的相互变换。用具有延时特性的神经元也可以构成串并行变换器。假定每一个神经元都在输入信号之后经过相同的延迟时间在输出端产生与输入相同的输出，把这些神经元一个接一个的串连起来，每一个神经元的输出接到下一个神经元的输入，就可以完成串并行变换的功能。第一个神经元输入串行模式之后经过若干个延迟时间，这一串神经元各自的输出端就可以得到并行模式。

利用神经网的模式联想功能，也可以产生或识别复杂的串行模式。例如有一个按时间顺序排列的模式串 $P_1, P_2 \cdots P_N$ 。可以在联想存储器中存入这样的联想关系： $P_1 \rightarrow P_2, P_2 \rightarrow P_3 \cdots P_{N-1} \rightarrow P_N$ ，把联想存储器的输出反馈到输入，此时只要输入 P_1 ，联想存储器就会联想出 P_2 ， P_2 再返回输入端联想出 P_3 ，从而按时间顺序联想出整个的模式串，如果模式串是首尾相接的，即 $P_N = P_1$ ，那么联想可以循环下去，直到发生控制中断。这样产生的模式串，可以作为运动的控制信息。同样的神经网络也可以用作模式串的认识，此时需要再加一个比较器，当输入口输入一个模式串时，每输入一个模式，都可以在神经网的输出口得到对下一个模式的预测，用预测模式与下一个输入模式在比较器中进行比较，得出差异，根据差异积累值的大小，可以判断输入模式串与存储模式串是否一致，实现识别的功能。

下面我们分析一个具体的神经网络模型——二多层神经网络，以

讨论回避指数爆炸的策略。二多层神经网络的模型示于图 4.4a。二多层神经网络是一个前馈网络，由输入层和输出层组成。与普通的二层神经网络不同的是，其输出层的神经元排有前后顺序，顺序靠后的神经元接受所有顺序在前的神经元的输出作为其输入，而后方的神经元对前方的神经元没有影响。其学习算法是这样的：假定其输入层有 N 个神经元，输出层有 M 个神经元，并按从 1 到 M 进行前后顺序的编号。学习时使用 L 个 N 维输入数组和与之对应

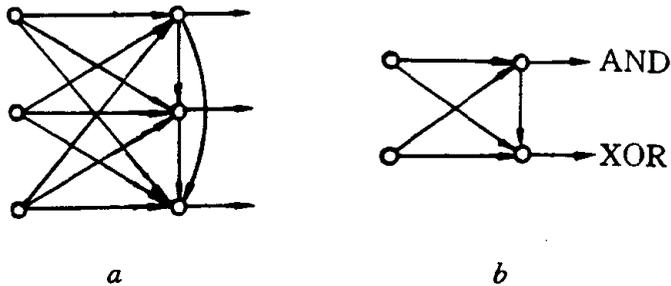


图 4.4

的 L 个 M 维输出数组作为学习样本。学习按照输出神经元的顺序逐个神经元进行，先对第一个输出神经元进行学习，选 M 维输出数组的第一个元素与之对应，学习时采用二层神经网络的学习算法。如果样本中的第一个输出元素与输入之间的关系不是线性可分的，那么学习失败，再依次选别的元素作为第一个输出神经元的学习样本，直到学习成功并记下第一个神经元对应哪一个输出元素。此后再对第 2、第 3 直至第 M 个输出神经元进行类似的学习，仍然采用二层神经网络的学习算法，所不同的是，前面已经学习过的输出神经元的输出同时作为输入参加学习。如果学习途中到某个神经元在未学习的输出元素中找不到一个与输入和前输出构成线性可分关系的元素，则学习由于失败而中止。图 4.4b 是一个可以学习成功的实例，4 个输入模式是 $(0, 0)$ $(0, 1)$ $(1, 0)$ $(1, 1)$ ，与之对应的 4 个输出模式是 $(0, 0)$ $(0, 1)$ $(0, 1)$ $(1, 0)$ 。

输出模式的第一个元素是逻辑与 (AND)，是线性可分的，因此作为第一个输出神经元可以学习成功，第二个元素是异或 (XOR)，对输入而言不是线性可分的，但是加上 AND 作为第二个输出神经元的第三个输入时，XOR 就成为线性可分的了，因此也可以学习成功。二多层神经网络由于总是使用二层神经网络的学习算法。因此学习收敛很快，不会遇到指数爆炸问题，但同时又是一个多层网，可以解决线性不可分的复杂非线性问题。当然二多层神经网络并不是在任何情况下都能顺利学习的，它还存在着两个难题。但是我们在这里讨论二多层神经网络并不是为了提出神经网的新的学习算法，而是为了说明一些重要的观点提供一个简单的模型。第一个难题是，我们怎样判断一个输出神经元的学习是否成功？从理论上讲，二层网只能解决线性可分的问题，因此如果一个输出神经元用二层网的学习算法学习线性不可分的问题则肯定失败。但是判断一个逻辑函数是否线性可分在理论上是一个指数爆炸的问题，这就是第一个难题的困难之所在。但是这个理论上的困难在实际上并不是太大的问题，如同计算机的操作系统原则上无法判断用户程序是否陷入了无限循环，但却可以对用户程序设定最大 CPU 时间，当用户程序运行超过这一时间时中断用户程序并通知用户。由于二层神经网络的学习算法对于线性可分的问题通常收敛比较快，所以可以规定一个不太多的学习次数限制，在限制次数之内学习不收敛者算作学习失败。第二问题是比较重要的，就是二多层神经网络学习的成功依赖于输出端的学习模式必须满足一种苛刻的条件，即至少有一个元素是线性可分的，而其他元素必须由更浅层的元素经过线性可分的步骤积累起来。在图 4.4b 的例子中，这个问题可以表现为我们怎么知道学习 XOR 问题需要 AND 作为前题条件呢？这个问题的回答只能是“不知道”，因为把线性不可分的函数化为多层线性可分的函数显然属于指数爆炸的问题。这个问题也可以用一般的多层前馈网模型来表现，如果可

以把一个复杂的线性不可分的问题分解成若干个线性可分问题的组合,就可以逐层地用双层网学习算法逐步完成多层网的学习,但是这个分解通常是困难的。

在充分认识上述困难的基础之上,我们可以得到认知科学中关于获得知识的一个重要的原理,我们称之为“短桥原理”。“短桥”是一个形象化的比喻。我们把知识看成是无知的大海中的岛屿,而认知即获得知识的过程,就相当于到达新的岛屿,增加可到达岛屿的数量。我们的认知能力受到指数爆炸的限制,不可能一下子走得很远,这就意味着我们所能建造的桥梁的跨度是有限的。我们只能建造短桥到达邻近的岛屿,这就是“短桥原理”。很多短桥连接起来,可以形成一个长桥,但是每次所能修建的只能是一个短桥。短桥原理是广义进化范式在认知科学中的表现,正如在进化过程之中,只能依靠小创造的积累来实现大创造一样,在认知过程中,任何新的知识都只能在充分接近的旧知识的基础上才能得到。对于任何认知系统来说,不是所有问题都是能够解决的,能够解决的问题局限于已有知识的有限距离之内。因此能够解决一切问题的“通用问题求解系统”(GPS)是不可能实现的。从这个观点出发来看二多层神经网络模型,就可以明白这个模型的意义不在于能否解决特定的具体问题,而在于它虽然只能解决线性可分的简单问题,但在每解决一个问题之后,它的知识都会增加,解决问题的能力增大,使得原来线性不可分的一些问题进入线性可分的范围。就如同图 4.4b 的例子在解决了 AND 问题之后使得 XOR 问题转化为线性可分的问题。这也是认知科学与传统人工智能在观点上的区别,我们建立一个认知系统主要不是关心它有多强的智能,而是关心它能否自己提高智能。冯·诺依曼在讨论计算机与人的不同点时指出,计算机的逻辑深度很深而人类思考的逻辑深度是很浅的。这是由于计算机所进行的运算通常是由人规定好而不需要学习和变更的,而人的有意识的思考则总是

伴随着学习过程，为了回避指数爆炸，不可能处理逻辑深度很深的问题，而在无需意识参与的固定化了的人脑信息处理过程之中，所处理的逻辑深度同样是很深的，例如视觉和听觉等信号的处理，是相当复杂而多重的过程。而在有意识的思考之中，例如要理解一个数学定理的证明，则要求证明的每一步都不能跳跃过大，否则就很难理解了。二多层神经网络也是如此，学习完成之后可以处理复杂的线性不可分的问题，但每次的学习都限制在线性可分的浅层逻辑范围之内。

二多层神经网络模型和“短桥原理”使我们注意到这样一个事实，即我们能否学习或创造某种知识受到我们过去所学到的或已经创造出的知识的制约。我们所拥有的浅层知识越丰富，学习或创造更深一层知识的能力也就越强，但是由于认知系统资源的限制，浅层知识不可能无限的扩展，因此对浅层知识的选择就成为至关重要的问题了，这种选择决定了认知系统未来的学习能力。但是由于通常未来的学习和创造的目标是未知的，而且即使知道也不能由深层信息的信息来判断出对浅层知识的需要，因而为知识的选择建立价值体系就显得非常重要，这就是我们特别重视价值体系的原因。人类之所以比其他动物远为丰富的感情世界，就是由于复杂的学习过程对感情的评价体系有着更高的需求。

4.3 神经网络群体

在软件工程学之中，模块化可以说已经成为一种标准的技术。这样作有多方面的原因，最主要的是人的思考资源是有限的，同时考虑太多的因素容易出错，所以对于大规模的软件需要将其分为众多的模块，一方面便于多人同时制作软件，也使得软件容易理解和维护，尽量简化模块之间的关联，以避免在修改程序时牵一发而动全身，每一个模块的功能应尽可能单纯，以便于制作和

调试。另一方面的原因是，功能单纯的模块便于在制作其他软件时重新组合而再次利用，形成软件资源的积累，提高软件制作的效率。实际上串行计算机的原理已经包含了某种程度的模块化保证，这就是CPU中的指令计数器，CPU在没有中断或转移指令的情况下是按次序从存储器中取出指令执行，这就意味着执行某一功能的多条指令在存储器里也是离得很近的，这样的程序列表比较容易理解。设想一台没有指令计数器的计算机，在每条指令后面带有下一条指令的地址，这样的计算机在原理上一样可以运行，指令可以在存储器中随机放置，但是这样的程序列表读起来的困难程度是可想而知的，而且还要浪费大量的存储空间来存储下一条指令的地址。

在高等动物的脑中，神经元的组织同样是模块化的，模块化的基本原则是相互紧密关联执行相同功能的神经元在几何空间分布上也聚在一起。大脑皮层是由柱状单元组成的，柱是由数千到一万个神经元组成的基本模块，人脑的神经元总数大约为一百亿到二百亿之间，也就是说共有一二百万个柱。每一个柱执行一种比较简单的功能，例如在视觉处理中对某个方向的线条作出反应。功能相似的柱再逐层聚成更大规模的模块，并与其上层和下层模块交换信息。每个神经元与其他神经元联系的神经纤维平均在数千至一万左右，与柱的规模大体一致，大多数的神经纤维（轴突或树突）是很短的，多数神经纤维与同一个柱内的其他神经元相联，少数与邻近的柱相联，与远处的其他模块相联的就更少了。假如人脑的结构不是模块化的，而是一个由一亿个神经元组成的均匀而巨大的神经网络，每一个神经元都与其他所有神经元相联系，那么神经纤维的数量要比现实的人脑增加一百万倍，由于神经元细胞核所占体积很少，人脑的尺寸主要取决于神经纤维的数量，那么人脑的体积将增加一百万倍，而且这样的神经网络在学习中一定会遇到指数爆炸的障碍。如果在这个均匀网中每个神经元的神经

纤维仍然只有几千个，而随机地和任意远的神经元联系，那么由于神经纤维平均长度增加，也会使大脑的体积增长成千上万倍，而且在其中找到可以组织有效功能的神经元的概率只有百万分之一，而绝大多数的连接是没有用的。这也就是在一个大规模系统中模块化的必然性。大脑是一个严格模块化的组织，并且是严密分工合作的复杂大系统。

复杂大系统中的模块可以按照不同的角度进行分类，形成级别和层次，以明确模块之间的关系。“级”是按模块规模进行的分类。若干具有类似功能的模块集成为一个整体，就构成上一级的大模块。例如在视觉处理过程中检测某一方向线条的柱是下级模块，若干个对不同方向起反应的柱组合起来，构成检测线条方向的上级模块，检测视野中各点线条方向的模块集合起来，构成检测方向场的更上级模块。“层”是按模块处理信息的深度来进行的分类。下层的模块是指比较靠近对认知对象的输入输出的模块，而上层模块则是相对远离传感器和执行机构的模块，下层模块处理的信息比较具体而局部，上层模块则处理较为抽象而整体的信息。以检测线条方向的模块为例，其下层模块是抽取边缘的模块，而上层模块则是从方向场中发现直线、曲线和角的模块。

在按模块结构组织起来的神经网络群体中，模块之间除了在上下层之间交换感觉与控制信息之外，还存在着协调问题。在普通的串行计算机之中，协调问题并不经常出现。一个模块化结构的程序，也同样有层次关系，上层模块可调用下层模块，但是由于CPU的串行动作，两个串行模块同时调用一个下层模块的现象是不会发生的。即使在多任务的操作系统中，几个程序可以“同时”运行，由于在更小的时间尺度上仍然是“分时”的，两个“同时”运行的程序可以调用相同模块而不发生冲突。只有在两个程序共用输出或输入装置，例如屏幕、键盘或打印机时，由于这些装置的时间尺度较大，才可能出现同时的冲突，需要采取避免

冲突的技术手段，例如分前台后台运行，分窗口，或打印信息暂存等。由于神经网络群体是完全并行工作的，因此与冲突及合作有关的协调问题就显得非常突出。

对于认知主体来说，模块之间的冲突主要是由两个“瓶颈”所引起的，一个是意识瓶颈，一个是执行机构瓶颈。从感觉器官来的感觉信息原则上可以并行处理而不会发生冲突，同一个感觉信息分到不同的模块抽取不同特征，不同的特征再会聚起来得到更抽象的信息，也不会发生冲突。但是由于意识资源的有限性，我们同时只能注意到感觉信息中的一小部分，这样就形成了意识瓶颈。当然感觉器官本身也不是完全没有瓶颈的，例如眼睛在一个瞬间只能注视一个方向，但这个瓶颈也是与意识瓶颈密切相关的。对于有意识控制的动作来说也存在着意识瓶颈，我们不可能同时专心致志地作两件事情。第二个瓶颈来源于执行机构不可能同时作两个不同的动作，就如同手不能同时既向前伸又向后摆。例如我们考虑嘴的三种动作：吃饭、说话和接吻，这三个动作不能同时进行，在一个短时间内只能进行其中的一个，如果要进行另一个，则必须放弃原来的动作。在这种情况下，控制这三种动作的三个模块之间就有可能发生冲突。基于上述原因，在一些模块之间就会产生冲突与合作的关系，一个任务需要若干模块合作完成，这些模块之间就构成合作关系，而这一任务与其他一些任务可能发生冲突，那么同时动作会发生冲突的模块之间就形成冲突关系。如果假定模块有工作和休息两种状态，那么为了保证由模块组成的系统协调地工作，需要让相互之间具有合作关系的模块同时处于工作状态，而让相互冲突的模块不处于同时工作的状态。如果认知系统的组织是“中央集权”方式的，那么协调可以由所谓的“中枢神经”统一加以控制。但是这种“中央集权”模型在理论上和实际上都行不通。从理论上讲，假定一个对人脑进行全面控制的中枢对于理解这个“中枢”毫无作用，只是把问题移了一个位

置而已，问题变成了中枢又由谁来控制，结果变成无限退行的问题。实际上对人脑进行神经生理学的研究也没有发现这样的“中枢”。因此这种模块之间的协调只能由分散自律系统模型来实现。在分散自律的系统之中，相互合作的模块之间具有“互兴奋”的联系，使得其中一方进入工作状态时促进另一方也进入工作状态；而在相互冲突的模块之间应具有“互抑制”的联系，使得一方进入工作状态时强迫另一方休息。在上下层之间，这种“互兴奋”和“互抑制”的关联可以靠感觉和控制信息来实现，而同层内各模块之间则可以靠专门的与同层模块相互联系的神经元的神经纤维来实现。模块之间的互兴奋和互抑制的关系显然可以按照 Hebb 学习准则逐步建立，在学习之前可以假定同层模块之间主要是互抑制的关系，任一时刻只允许少数模块同时工作。在相互冲突的模块之间，存在着竞争关系。在哪种情况下应该由哪个模块争得工作状态，需要一个统一的评价准则，这里也体现出感情评价体系在系统协调中的重要作用。

有一种理论认为，人的大脑实际上只使用了不到五分之一，五分之四的潜力还没有被开发，如果把这五分之四的潜力开发出来，人可以变得聪明得多，比方说智商可以达到五百。首先从进化的角度来说，如此不经济的事情是很难想象的，再者也没有见到过这种开发成功的实例。而且如何计算人脑的“潜力”也是一个困难的问题，如果这个理论的根据是人脑只有一小部分模块处于工作状态的话，那么结论显然是错误的。如果把所有模块“开发”成可以同时工作的话，就完全破坏了整个系统的协调。此外由于人在一生中不断学习的需要，大脑中必须准备一定的多余的空白模块，否则人就进入一种不可能学习任何新技能的状态。按照维纳的说法，我们的长期记忆资源可能是有限而不可再生的，如同《驴皮记》中的那张驴皮，只会缩小不会长大，那么一下子用完这块驴皮当然是不明智的，这意味着失去学习和记忆的能力。因此

一个百分之百的“潜力”都被开发出来的人，只能是完全失去协调并且毫无学习能力的人，其智商不是五百而会接近于零。

本节只讨论了认知主体作为具有模块结构的神经网络群体在组织及模块之间关系方面的一些基本原则，没有涉及人脑具体的总体结构和模块内部的神经网络结构，但是这些原则可以为具体结构的设计提供必要的基础。目前神经生理学的研究对于人脑较低层次的信息处理过程已经取得了很大进展，但是对于较高的接近意识的层次还知之甚少。其主要原因并不是缺乏必要的实验手段，人脑高层次活动与低层次活动所需的检测手段并没有太大的区别，主要的区别是对于低层次的活动我们已经掌握了很多关于信息处理的理论和模型，可以利用这些理论和模型提出各种假说，有了假说之后，才能设计必要的实验对假说进行验证。而对于人脑的高层次神经活动，至今还没有一个较为全面的能够自圆其说的理论，因此尽管可以通过实验得到一些零散的现象，但是却难以对这些现象给出总体的解释和说明。基于这种现实，目前我们的首要任务是对于更高层次的神经活动建立起关于功能和机制的理论模型，只有在这个基础上，对于认知系统的较高层次才有可能进行有效的神经生理学的研究或是设计人工认知系统的结构，在计算机软件工程之中，也总是功能设计先于结构设计的。因此在本章之中对于认知系统的分析主要侧重于功能的分析，讨论各种功能之间的关系。为了对功能进行说明和解释，会采用一些结构性的模型，但是采用这些模型既不意味着模型是唯一的，也不意味着是最佳的，更不意味着人脑的结构与这些模型完全一致，而仅仅意味着我们所讨论的功能在工程上是可实现的。对于这些功能，我们所关心的不仅是功能本身的机理，更重要的是通过进化与学习产生这些功能的过程，也就是说更重视动态的过程。

4.4 从动物到人的遗传与学习

在认知科学的研究之中，我们把人脑的认知作为认知系统的典型代表，但是研究动物的认知过程，对于理解人的认知是有很大的帮助的，甚至是必不可少的。按照短桥原理，关于动物认知的知识可以为我们准备离人脑认知岛屿比较近的小岛。这就象我们看到一个机械产品，对于缺乏工业常识的人来说，想出这个产品的制造过程几乎是不可能的，但是如果我们已经了解很多基本的加工过程，例如铸造、锻压、冲压、焊接、切削、注塑、吹塑等各种加工工艺，就可以很容易地编出这个产品的全部制造工序。研究动物行为的有利条件是，动物的本能行为和学习过程都比人要简单得多，使我们比较容易找出一些基本的低层次的原理，由于动物的种类很多，复杂程度也不同，使我们有可能从低到高逐步地理解这些原理的进化过程。而且由于人是由动物进化而来的，而进化又是一个积累过程，所以我们有理由认为动物的一些认知机理在人的认知过程之中仍然起作用。通过对动物认知的理解，不仅可以从人与动物的相似之处使人的认知更容易理解，而且可以从人与动物的相异之处使这种理解更加深化。关于低层次认知的知识不仅为组建高层次认知提供基础，而且也能突出高层次认知比起低层次认知的超越之处。

与人相比较，动物的行为之中本能的成分较多而学习的能力较差。本能与学习的最大区别在于，本能是生物个体通过遗传基因获得的行为能力或行为方式，而学习则是通过后天的学习得到行为能力或行为方式。本能是遗传的而学习的结果是不遗传的。但是要区分本能与学习并不总是很容易的，二者之间很有些纠缠不清的关系。首先，我们不能简单地根据出生时不具备某种能力就判断这种能力不是本能而是学习得到的，因为生物的发育在出生

后并未停止，遗传基因里的程序可能到老都还在起作用。我们需要通过一种行为方式与遗传的关系，例如在同一个属中的不同种之间具有类似或相同的行为方式来区别本能和学习，有时也可以采用对生物隔离学习条件来判断一种行为方式是本能还是学习结果。本能与学习之间往往有密不可分的关系，对于任何一种学习来说，都必须以本能作为基础，本能为学习提供低层次的信息处理，提供学习算法，提供方向性的规定（评价准则）。按照短桥原理，任何学习都不可能从白纸开始，需要一系列的先决条件，这些条件在一开始只能由本能提供，其后有一部分可能由前期的学习提供，但本能的部分仍是不可缺少的，本能决定了大部分的学习机理，而这种机理是比较稳定而深层的，不象学习内容那样容易发生变化，因此对于学习的研究中与本能有关的部分更值得重视。反过来说，有些明显属于本能的行爲，在发育过程中却离不开学习，例如猫捉老鼠的动作，在猫科动物中基本上是类似的，可以说是一种本能，但是猫在成长过程中总是不断用游戏的方式来练习这个动作，这是由于跳扑动作的准确与猫的动力学特性例如质量和肌肉弹跳力有密切关系，而猫的胖瘦却是基因所无法完全规定的，在这种情况下基本行为方式虽然是遗传的，但其中的部分参数却需要通过学习进行修正。人的直立行走也是一种本能，但是也需要一定程度的学习才能实现这种本能，如果失去学习的机会，人也会象“狼孩”那样只会爬行。

为了探讨本能的一些性质，我们观察一种简单的单细胞生物草履虫的运动方式。生物学家对草履虫看起来毫无规律的运动进行了详细的研究，把草履虫的运动总结成几条规则。草履虫在正常情况下作直线运动，食物越丰富则速度越慢，当它碰到障碍物时，先后退一段距离，再随机地转一个方向，然后继续作直线的前进运动。这几条规则极为简单，以至于对于计算机的初学者来说，编一个很短的程序，在计算机屏幕上模拟草履虫的运动也是

不困难的。从进化的似目的论的角度来看，这几条规则使草履虫成为一个“适者”，缺少其中任何一条都会使草履虫的适应度下降。第一条作直线运动是非常重要的，如果运动不是直线而是有一个固定的曲率，那么草履虫会在没有障碍的情况下作圆周运动，当圆周轨迹上的食物吃光之后，草履虫只好饿死了。草履虫保证其直线运动的机制也是很有趣的，草履虫前进的同时作着自旋的动作，产生自旋的原因估计是鞭毛的旋转运动的反作用力。如果没有自旋，草履虫体型极小的不对称性也会引起前进的偏向，形成圆周运动，而草履虫又确实是不对称的，口长在身体的一侧，自旋自动地抵消了不对称的偏转，使草履虫作直线运动。食物越丰富速度越慢则使草履虫较长时间呆在食物丰富的地区。假设在计算机屏幕上模拟很多草履虫，如果没有这条规则，可以看到草履虫在屏幕上分布比较均匀，而加上这条规则，再在屏幕上设定一些食物丰富的地区，就会看到在食物丰富的地方草履虫密度较大，就如同堵车的地方车比较多一样。在碰到障碍时如果不后退，就地改一个方向再前进的话，有很大的概率是再碰在同一个障碍物上，如果后退之后不是随机地而是旋转一个固定角度，那么在某些特定的环境中草履虫的运动轨迹也有可能陷入无限循环而无法逃脱。在生物的本能行为、学习乃至进化过程中，随机的变化或变异经常是需要的，我们通常对这种随机信息的来源不作深入探讨，这种随机信息到底是真的随机还是象计算机里制造的那种伪随机，还是由混沌现象产生的象随机的东西，都无所谓，我们更注意选择。但是混沌理论无疑为我们提供了一个产生随机信息的容易实现的简便方法，使得在生物中随机信息的产生有了更令人信服的根据。

从进化的眼光看，本能具有适应的性质是理所当然的，随着不适应的个体被淘汰，产生不适应本能的基因也被淘汰，因此任何一种本能，包括支持学习的本能，都必须是适应性的。它们不

应使持有这种本能的生物成为不适者。换句话说，本能与生物的其他性状一样，是源于进化的。本能的另一个性质是其机械性，这一点并不象适应性那么一目了然。草履虫的运动是相当机械的，按照上述的几条规则我们可以制作与象草履虫那样漫游的机械玩具。但是草履虫毕竟太简单了，更复杂的动物本能往往显得不那么机械，但是如果考虑到本能必需由基因来决定的话，其复杂性显然是受某种限制的。我们平时观察到的象蜜蜂筑巢或蜘蛛结网这种精妙的本能，经过生物学家之研究和分析，往往会被简化成有限的几条规则，以至于我们情不自禁地感到有相同能力的机器并不那么难作，难怪笛卡尔要把动物看成是自动机器，这不仅有形而上学的理由，可能也包括敏锐的直觉。生物本能的另一个性质是其主动性。在草履虫这样简单的例子中，也可以看到一种不撞南墙不回头的劲头，草履虫首先要运动，才可能碰上障碍物，这与常见的刺激——反应模型（ $S-R$ 模型）有点不一样，生物通常不是坐等刺激来临，而是主动地寻求刺激，这一点是在构筑或分析认知系统时不能忽视的。

草履虫能不能学习，这一点在生物学界可能有不同的看法，但我们前面所讨论的模型是不包含学习过程的，下面我们看一个蜜蜂学习颜色的简单例子。生物学家对蜜蜂的辨色能力进行过详细的研究，基本的实验方法是，在一个桌子上放上盛着糖水的透明盘，通过改变照明或色纸的方法可以任意设定盘的颜色，在蜜蜂吸过糖水之后拿掉糖水，看蜜蜂第二次来时是否还落在原来的颜色上。为了排除其他因素例如位置、气味等影响，实验比上面说的要复杂一些。实验的结果是，蜜蜂确实有辨色的能力，比较容易学习的是黄色和紫色，对绿色和紫外线不敏感；学习发生在吸蜜之前的 0.5~3 秒之间；学习之后的记忆时间随吸蜜次数增加而增加，一次吸蜜的记忆时间约为一天，三次吸蜜之后可以记忆十几天以上。这种学习方法的主动性是显而易见的，蜜蜂如果不出

去采蜜学习是无法进行的。适应性也比较明显，对绿色不敏感是由于叶子都是绿色的，相当于背景，而紫外线是天空的颜色，蜜蜂可以看到紫外线，但主要用于靠太阳的位置进行长距离导航，而对花的颜色的反应则用于近距离寻找目标，不应混淆。对于吸蜜之前 0.5 秒至 3 秒的颜色进行记忆是由于这一段时间恰恰是蜜蜂进行寻蜜的短距导航和着陆的时机，也就是说把颜色与使用的场合 (*Context*) 信息正确地结合起来。由此可知，对于有效的学习来说，必须由本能提供严格的学习范围，学习时机和学习方法。这里也体现了短桥原理。

关于蜜蜂对于颜色的学习似乎也可以找到非常机械的学习算法，这就是在吸到蜜时记住 0.5 秒至 3 秒之前所看到的颜色，并在下次吸蜜之前用这个颜色作为导航目标。为此需要一个短时的寄存器，眼睛所看到的颜色经过 0.5 秒的时间延迟送到寄存器去，在寄存器里过 2.5 秒自动消失。一旦吸到蜜时，把寄存器里的色信号转移到长期存储器中去，下次采蜜时随时比较已经记住的色信号与看到的色点，当二者一致时，把色点作为导航标志飞向这个色点。由于蜜蜂是四面八方都看得见的，还需要规定所学的颜色是前下方看到的颜色。上述这种算法对于蜜蜂的辨色学习来说实际上是不够的，它无法解释为什么吸蜜次数多了记忆时间就长。这种算法更接近另一种学习，也就是生物学中所熟知的“印痕” (*Imprinting*) 学习。很多动物在出生后不久，就会记住自己的母亲，并且凭着这种记忆随时跟随自己的母亲行动。这种学习是不靠赏罚的，不需要吸蜜这种外部刺激，而是完全靠时间程序的，即出生后的某一时间区间所看到的运动物体，将被记住作为追随行动的目标。生物学家用比较严格的实验方法测定出野鸭的印痕学习期是孵化之后 13~16 小时。印痕学习的记忆是终生不忘的，所记的知识不一定象追随运动那样当时使用。例如对于求偶对象的识别是靠出生不久的记忆，而这个识别的知识要到性成熟之后才

被使用。象蛙鱼或鲟鲤鱼这类在河流上游产卵而在大海中长大的鱼也是靠印痕的学习方法记住“祖河”的气味，待长成之后凭这种气味（水味？）再回到“祖河”里产卵。这种学习本能当然也是适应的，如果误入其他河流，可能找不到合适的产卵水域，或者由于误入者太少而找不到配偶，而在“祖河”里不会遇到这种困难。印痕学习也需要本能的支持，由本能规定记忆的时间、目标选择所依据的特征、要记忆的特征、起用记忆的时机和方法等等。因此任何有成效的学习都必须立足于足够的“先验知识”之上。

为了更合理地解释蜜蜂对颜色的学习过程，我们需要一个更完善的学习模型，这就是进化稳定学习（ES 学习）模型。这个模型最基本的原则是，学习的结果应该使生物获得最大的利益（适应度），当生物具有若干种可选择的行为方式（战略）时，对某一种战略选择的概率应该等于历来采用这种战略所获得的利益在总利益中所占的比例。在这种情况下，选择任一种战略所获得的利益是相等的。ES 学习的基本算法就是提高每次获利较多的战略的频度也就是提高有利行为的概率。下面就是根据文献（Harley, 1981）改写的学习算法：

$$S_i(t+1) = mS_i(t) + (1-m)r_i + Q_i \quad (0 < m < 1)$$

$$P_i = \frac{S_i}{\sum_j S_j}$$

式中的 S_i 是对应于第 i 种战略的利益积累量。 m 是记忆常数， m 较大表示对过去所学的知识不易遗忘，但是随环境变化而改变行为方式较慢。余差 r_i 表示动物的本能倾向，当学习不起作用时，系统恢复到由 r_i 决定的概率分布。 Q_i 是当使用 i 战略时获得的利益。行为总要付出能量，当收益少于支出时， Q_i 是个负值。 P_i 是采用第 i 种战略的概率，等于战略 i 的利益积累量与总利益积累量之比。学习的过程是这样的，每次行动都按照各战略的概率分布 P 随机地决定采用一种战略 i ，然后在行为完成之后评价出该

行动的收益 Q_i ，按上述式子计算出 $S_i(t+1)$ ，对于未被采用的 $S_j(t+1)$ 也按相同式子计算，只是令 $Q_j=0$ ($j \neq i$)。如果 Q_i 大于其他 Q_j ，则 P_i 将会逐步增加，即动物的学习总是使收益较大的行为增加，直到达到平衡。但是由于 r_j 的存在且 $m < 1$ ，即使在 Q_i 总是大于其他 Q_j 的情况下， P_i 也不会达到 1，也就是说，其他 S_j 不会变成 0。因为如果其他的 S_j 为 0，会使 P_j 也为 0，这意味着生物只按成功的行为方式行动，不再试探其他行为方式，那么当环境变化时，动物就不可能通过学习改变行为方式以适应新环境。*ES* 学习的动态过程与进化中的选择过程极为相似，策略 i 对应于某种基因，收益 Q_i 相当于基因 i 的性状的适应度， P_i 则相当于基因 i 在群体中所占的比例， r 的作用则体现多样化的价值。学习与进化的重要的不同之处在于，进化所依据的评价准则是定义明确而唯一的适应度，而收益 Q 却需要由“主观”的评价准则来确定，于是如何决定评价准则就成为一个重要问题。但是从进化的观点看，评价准则本身受着进化的考验，如果评价准则导致动物采取降低适应度的行为，那么这种评价准则会在进化的过程中被淘汰，也就是说在进化过程中产生的评价准则一定是适应的。

ES 学习的一个典型的例子是 Milinsky (1979) 所作的刺鱼觅食实验。在一个长水槽中放 6 条鱼，水槽的两端有喂食口，左端的投食量是右端的二倍。鱼刚刚进入水槽时自由地游来游去，基本上在水槽中均匀分布，但是经过一段时间的学习之后，鱼在水槽中的分布渐渐接近于 *ES* 学习的平衡状态，左边 4 条，右边 2 条。当然并不是每一条鱼都以 $2/3$ 和 $1/3$ 的概率呆在左边和右边，有些鱼更多地呆在左边，而同时必有一些鱼更多地呆在右边，只保持集团平均成最佳分布，但没有一条鱼是只呆在左边而不去右边的。如果把左端和右端喂食的量反转过来，经过一段时间的学习之后，鱼的分布又会变成左边 2 条，右边 4 条。用计算机按照 *ES* 学习算法对鱼的觅食行动进行模拟的结果与实验中鱼的行为过程非常一

致。

用 *ES* 学习模型来解释蜜蜂学习颜色的过程，更符合实验观察的实际结果。假定蜜蜂能识别红、黄、兰、紫四种颜色，四种颜色在学习之前平等地吸引蜜蜂，蜜蜂在寻蜜时随机地采用不同颜色作为引导色，当采用某种引导色时吸到了蜜，该色作为引导色的概率上升。这与实验的结果一致，蜜蜂在吸过蜜之后并不是百分之百再回到同一种颜色，而是随着吸蜜成功的次数增加，回到同一种颜色的概率逐步提高，相当于 P_i 的增加。在假定忘却速度一定的条件下，吸蜜成功次数多的蜜蜂 S_i 也比较大，记忆时间也会更长。如果某种颜色的花被大多数蜜蜂选择吸蜜，其存蜜量就会下降，蜜蜂对该色花的吸蜜利益下降，导致蜜蜂更多地试探其他颜色，发现新的蜜源。

Dawkins 为了说明 *DSS* (*Developmentally stable strategy*) 的想法举了一个巧妙的例子，对于进化稳定学习很有说服力。这是 Baldwin & Mees (1979) 所作的一个实验。在一个长猪圈里关两头猪，一个强者和一个弱者。猪圈装有机装置，在猪圈的一头踩一个踏板，另一头的喂食口就会流出一定量的猪食。经过一段时间的学习之后，猪掌握了取食的操作方法，形成了稳定的行为方式，每次总是强猪去踩踏板，弱猪在喂食口等着，强猪踩了踏板之后奔到喂食口把弱猪赶开，吃弱猪吃剩的猪食。这个行为方式成为稳定解是显而易见的，强猪踩了踏板得到吃剩猪食的利益，而弱猪等在喂食口则可以在被赶走之前吃到猪食，也是有利益的。但是相反的行为方式却不稳定，如果是弱猪去踩踏板而强猪等在喂食口，由于弱猪赶不走强猪，猪食全被强猪吃光，弱猪得不到任何利益白费了力气，所以弱猪去踩踏板是不被增加的行为。即使每次出的猪食很少，在强猪赶到之前大部分被弱猪吃掉，给强猪只剩很少的一点，强猪踩踏板的行为仍然是稳定的，有一点吃的总比没有强。可以想象，长此以往，弱猪由于吃得较多而变壮，

强猪由于吃得少又奔波劳累而日渐消瘦，总有一天力量对比发生逆转，强弱身份互换，原来的强猪变成弱猪，才能开始享受“弱者的利益”。由于野猪是一种单独行动的非合作型动物，而上述环境又是自然界所罕见的，所以猪不大可能通过学习发现合作的方法，例如轮流踩板和吃食。如果猪世世代代生活在这种猪圈里而又能实现自然选择，那么几万年或几十万年之后，也可能会进化出一种具有绅士风度的猪，吃食时只吃三分之二，剩三分之一而不是全部吃光，那么强者就可以永远保持其强者地位，不会让弱者反而占便宜。如果圈里关的不是猪，而是狗或狼这种合作型动物，合作体制可能很快会通过学习建立起来，因为这些动物不象猪那样习惯于独吞。

ES 学习准则不仅对动物具有重要意义，对人来说也是一种主要的学习方法。例如一个可以自由移动摊位的小商贩，即使对市场分析一窍不通，也可以凭每天的销售利润选出最佳的摊位，而大量的商贩通过这种选择可以形成最佳摊位分布。选择摊位的过程比较复杂，其中也涉及多种约束和有意识的思考过程，而 *ES* 学习是一种基础性的低层次学习方式，并不一定需要意识的参与，完全可以自动进行。我曾经遇到过这样一件事：在一次教授与学生的联欢会上，学生问几位教授，三点水加一个来字念什么，思考片刻之后一位教授回答：“念来”学生们说“对啦”，接着又问，三点水加一个去字呢？想了一会儿一位教授说：“大概还念去吧”，学生们哄堂大笑。这种错误可以用 *ES* 学习模型来解释。中国人在别人问某个字的念法时通常用两种方法解决，第一种是通过形象思维先在脑子里构成这个字，然后利用模式识别的手段直接联想出读音，第二种方法是遇到生僻字不知道读音时读其右半边作为尝试性答案。由于“涑”字是不常用的字，第一种方法无效而第二种方法成功，导致第二种方法的采用概率提高，以至于“法”字本是常见字，却用了第二种方法造成错误。加上来与去在语义上

的对称性，以及教授通常接触的人不会问三点水加个去字念什么的问题，这个问题本身显得很生僻，更加大了错答的可能性，但是第一个问题的诱导作用是最主要因素，如果不问第一个问题而直接问第二个问题，大多数人是不会出错的。这个例子说明 *ES* 学习规则在无意识之中操纵着我们的思考，而错误与学习是密切相关的。没有学习功能的计算机是很少出错的，而善于学习的人则常常出错，因为学习的最基本的方法是试错法，与进化的变异与选择一样，变异就是遗传的错误，而无变异的生物是不可能进化的。*ES* 学习总要使利益很小的行为概率不为零，也就是以错误行为为代价，换取学习的可能性。可以说，错误是学习的必要条件。

ES 学习规则是动物学习的基本而重要的规则，但并不是全部。*ES* 学习规则所规定的是如何在若干种行为中选择最佳的行为方式，但在选择之前需要事先存在一些行为方式以供选择。对于选择来说，客观条件与主观需求无疑是重要的因素，因而对客观事物的识别也是非常重要的。无论是本能性的行为还是包含学习的行为，都必定需要对于行为对象和条件的识别和适当的操作方式，识别能力与操作方式有些是完全靠遗传而来的本能，有些（对于人来说大多数）是通过后天的学习得到的。在动物行为的最简单的刺激反应（*S-R*）模型之中，识别与操作分别对应着刺激 *S* 与反应 *R*，因此关于识别的学习我们简称为 *S* 型学习，关于操作的学习称为 *R* 型学习。*S* 型学习与 *R* 型学习是密切相关的，但把二者区别开来讨论也有助于研究二者的关系。在人工智能领域，*S* 型学习得到的知识是关于“是什么”的知识，也可以叫作宣言型的知识，而 *R* 型学习得到的是关于“怎样作”的知识，或称为手续型的知识。对于计算机来说，识别比操作控制要困难得多，制造出能打印汉字的机器比较容易，而制造能识别汉字特别是手写汉字的机器则很难与人的识别能力相比。而对于人来说事情似乎相反，认识人几乎不需要学，学认字比学写字容易。其实因为识别相当

于数学上的反问题,对于动物来说同样是困难的。人与动物之所以表现出高度的识别能力,是由于人脑在资源安排上付出了很大的代价,人的大脑皮层中用于识别的面积比用于控制运动的面积大很多倍。

动物的生存依赖于合乎生存目的的行为,因此在动物的神经结构中通常包括很多先天的识别和操作的本能。例如青蛙对于一定距离的水平运动目标可以引发捕食动作。象图 4.5 所示的形状按箭头方向移动可以引发水禽的惊恐、逃避等反应,即使是从未见过鹰的雏鸟也不例外,但是如果按相反方向运动则不引起反应。乌鸦见到带有柔软黑色物体的移动物会引起攻击反应,可以想象一只狐狸叼着一只死乌鸦会被其他乌鸦群起而攻之,而人拿着一块黑抹布走动,也会引起乌鸦相同的反应。按照动物行为学家洛伦兹 (Von Konrad Lorenz)

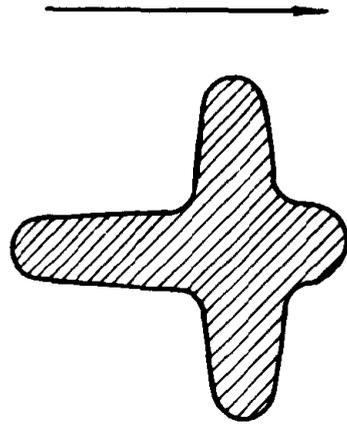


图 4.5

所提出的模型,动物并不是象 $S-R$ 模型所表示的那样简单地对外部刺激产生“反射”式的反应。在动物的中枢神经系统中,对应着每种行为有一个中枢兴奋机构 (*Central Excitatory Mechanism, CEM*),在 *CEM* 中按照动物的需要生产和积蓄冲动兴奋,当受到该种行动的特定的键刺激时,积蓄的冲动释放出来,驱动有关的运动中枢执行特定的程序。如果在脑中的该中枢部位插入电极用电脉冲进行刺激,也可以引起相应的行动。当冲动积蓄较多时,动物会实行“欲求行动”,主动寻找键刺激。*CEM* 模型比 $S-R$ 模型要完善一些,可以解释动物行动的主动性和目的性。二者的共同点在于都需要键刺激作为行为的起因。

键刺激这个词顾名思义,是一种类似于钥匙的刺激,就象一把

钥匙开一把锁一样,对于某种特定的行动需要某种特定的键刺激来引发。键刺激有的比较单纯,有些则相当复杂。动物对键刺激的认识机制;涉及模式识别这一重要的领域,而识别能力的 S 学习过程与认知密切相关。动物所受的刺激总是来源于某种感觉器官,因此键刺激识别所需要的信息处理也随感觉种类的不同而具有不同的原理。嗅觉可以说是动物最古老的感觉,在人脑中处理嗅觉的部分是进化早期的部分。嗅觉的优点在于特异性强而信息处理比较简单,嗅觉的传感器可以对某种特定的生物化学产物起作用,由于生物化学产物的来源非常特别,很难遇到干扰,因此灵敏度可以非常之高,由于作用因素的特异性,在信息处理方面就非常简单,可以不经处理直接用于识别,这对于头脑比较简单的生物是非常有利的。嗅觉的缺点是作用距离较小,且易受气流影响,定向性差,从原理上一个嗅觉传感器是无法断定方向的,要加上空间分布的多个传感器,或是依靠风向感觉,或加上动物的移动,才能测出气味浓度场分布,找到气味源的方向。听觉信息的定向性较好,可以非直线传播,作用距离较远,但特异性稍差,信息处理比嗅觉要复杂得多。视觉的定向性极好,特异性也很强,作用距离远,但由于信息量大,用于识别的信息处理极为复杂。从维数上看,嗅觉是 0 维信息,听觉是 1 维而视觉是 2 维,如果加上时间轴则分别为 1, 2, 3 维,由此可知信息处理的计算量是视觉大于听觉,听觉大于嗅觉,人脑中神经资源的分配也符合这个关系。当然由于各种动物对于信息的要求不同,其处理的方法也不同,神经资源分配也与人有很大差异。

键刺激往往是复杂的模式,正如钥匙不只有一个齿。一种气味的感觉可能由不只一种分子的检测所组成,一种声音可以分为音调、音色、节奏、旋律、强度、方向(相位差)等多重因素,视觉的成分则还要复杂得多,一种键刺激又可能包含各种感觉的因素。关于一种键刺激中所包含的各种感觉成分之间的关系,生物

学家们进行过多种研究，得到了“异质叠加定律”，即一个键刺激如果由多种感觉成分组成，那么键刺激强度等于各成分单独刺激强度之和，是一种简单的算术关系。例如，一种河雀科的雄鱼在遇见同种的其他雄鱼时，会发生变色反应，根据变色的程度可以测出受刺激的大小（实际的实验要复杂一些）。在实验中发现，雄鱼并不是对目标的整体效果（格式塔）起反应，而是分别对各个特征起反应，只给雄鱼看另一条鱼的单个特征，例如身体的兰色光泽、眼部的黑色斜线、颈下的黑色突起、胸鳍上的黑边、摆动等都具有键刺激的效果，而整体效果等于部分效果之和。还有人研究了蜜蜂分巢，选择新巢时对地点的总评价也是各评价因素之和。在生物学领域得出这种定量的结果是很不容易的，但从信息处理角度看，一个神经元求加权和的计算就足以完成这种功能，求和只是求加权和的特例。这样一种运算显然是线性可分的，如果在加权和之后加上一个阈值函数就构成模式空间中的一个超平面，如果阈值取得较大，可以使得各个特征缺一不可，任何单独特征都不构成刺激，而只当全部特征存在时才构成刺激，有点总体大于部分之和的味道，但这只是“伪格式塔”，相当于逻辑与门，真正的格式塔还要加上各特征之间的关系。但是伪格式塔可以成为真格式塔的部分信息，这是一种分析的态度。格式塔学派的最大缺陷在于拒绝分析的态度，把格式塔看成是不可分析的整体，成为一种不可理解的存在，使理论失去了明晰性。当然，不能轻视格式塔的意义，特征有很多都是有形的，例如鱼的黑色斜线、突起、黑边，也都是格式塔，图 4.5 的形状也是格式塔，问题在于弄清格式塔与非格式塔和伪格式塔之间的关系，才能使格式塔成为一种有用的概念。心理学家的实验表明，婴儿在某个时期对不按规定位置散乱分布着眼睛、鼻子、嘴的图画起反应，但过一段时间则对此种图画不再注意，只注意五官端正的图形。这个实验暗示在形成格式塔感觉的过程中也有非格式塔的阶段，格式塔可

以由非格式塔组合而成。

对键刺激的区分，是模式识别的过程。在模式识别领域，通常把识别分为特征抽取和判别两个步骤。这种分法如果只从模式识别的角度上看具有人为的任意性，但是如果同时考虑通过学习提高模式识别能力的过程，或者说模式认知过程，这种区分就有一定的必然性了。识别 (*Recognition*) 就是再认知 (*Re-cognition*)，获得或提高模式识别能力的学习过程就是模式认知过程，只有理解了模式认知过程之后，才能更深刻地理解模式识别的机制，*S* 型学习过程就是模式认知过程。根据学习的短桥原理，我们需要把学习限制在回避指数爆炸的可行范围之内，即线性可分的范围之内，从这个前题出发，可以规定判别函数必须是线性的，而学习只发生在判别函数的参数调整之中，在神经网络中则对应着二层神经网络的学习。因此我们把模式识别分为特征抽取和判别两步时，就可以根据这个原则，把最终的线性可分的信息处理过程称为判别，而把前期的复杂信息处理叫作特征抽取。当然，一个学成的模式识别能力的识别结果也可以看作一个特征，成为上一层学习的基础，这就构成了可以逐层建构的软硬结构，由多个短桥组成长桥。特征抽取是模式认知的硬结构，判别则是其软结构，判别的学习一定要在特征抽取的基础上进行，而学习的结果又增加了特征抽取的能力，从而使特征抽取能力在多次的学习中积累提高。

印痕学习是比较简单的 *S* 型学习过程，通过对这种学习的分析可以使我们了解 *S* 型学习所需要的前题条件。第一个条件是由短桥原理导致的，即对于要学习的模式需要已有的特征抽取能力作为学习的基础，待学特征的数量不少于学习后选择特征的数量。第二个条件是学习所得的键刺激的用途应是已知的，并在相应的场合背景下进行。第三个条件是在学习之前必须具备识别学习目标的能力。第四个条件是对所有学习通用的，即已具备学习的算

法和学习的动力。对于印痕学习来说，可以认为这些学习条件都是由本能提供的。对于野鸭类的雏鸟追随母鸟的印痕学习来说，学习的动力来自雏鸟的发育程序，其他条件则依赖于雏鸟在该发育阶段已具备的本能，雏鸟对于应该追的对象已有先验的知识，即能够把运动的、出声的对象判断为追随对象。第三个条件可以作为第一个条件的一部分，但第一个条件应准备更多的特征，例如颜色、形状、音调、音色等等。第二个条件是在追随过程中学习用于追随的键刺激。在追随过程中，母鸟的全局颜色、局部颜色、花纹形状、音色特点、音调特点、动作特点等逐步被记忆，最终达到可以与其他雌鸟相区别的程度。

比印痕学习更复杂的 S 学习过程同样需要上述的四个条件。乌鸦是鸟类中学习能力强的一种。乌鸦学习的动力在于其好奇心，一般学习能力强的动物其好奇心也强。乌鸦的学习过程也有固定的学习算法，看起来程序性很强。当乌鸦发现未知的动物时，首先把它看成敌害，从它的背后发动攻击，如果遭到强烈抵抗，则用对付敌害的方式行动，如果目标力量不强，则把它看成食物，杀死之后试着吃掉，如果味道不符合食物的要求，则看成其他物品，以后作为掩盖藏食之用。这个学习过程需要事先能够识别动物与其他目标，识别动物头尾前后的区别，否则无法从后方攻击。需要用已知特征把目标与已学习过的动物区别开。在键刺激的用途方面对目标作了三种假定，即敌害、食物、杂物，根据学习的结果确定哪一个假定是正确的。模式识别的过程是一个归类过程，但是按照丑小鸭定理，对事物进行“客观的”分类是不可能的，乌鸦学习中体现了分类的一种“主观的”原则，即按照目标的用途及处理方法来分类，这样就可以使学到的键刺激起到提高适应度的作用。如果所有动物认知都遵循这一分类原则，这个分类就具有一定的客观性了。根据适应度在进化中的普遍意义，这种分类准则对于模式认知来说也是具有普遍意义的。

在模式识别领域，特征抽取是一个困难的问题，按照前文的观点，特征抽取应该满足的一个条件是使判别成为线性可分的。但是这个条件对于决定抽取什么特征和怎样抽取特征来说是很不够的。我们常常希望有一种从识别的任务决定出特征的选择和抽取算法的普遍方法，但是这种希望显然是不可能实现的，这就如同由某一种生物的结构去推算其进化或发育过程一样是指指数爆炸的。所幸的是我们对于进化的机制并非一无所知，关于进化的知识有助于我们进一步探讨特征抽取的一些原则。进化的机制通常不能告诉我们哪一种性状会在进化中产生出来，但可以说明某一种性状为什么未被淘汰，而被保存下来。特征抽取作为感觉信息的处理方法是在与感觉器官的结构与功能相互适应的协同进化之中形成的，它必然受到进化过程中适应度的约束。这个机制对于说明象青蛙这种主要靠本能生存的动物的特征抽取能力是有用的，但是对于象人这样具有高度学习能力的动物来说就不够具体了，由于学习是一个开放的过程，目标选择的自由度很大，需要什么样的特征抽取能力从适应度是难以直接评价的，此时需要一些间接的评价准则，例如稳定性、经济性和可预测性。我们对于物体颜色的感觉是由光源光谱特性和目标的反射特性共同决定的，但是光源光谱特性的变化对我们生存的意义并不大，而目标的反射特性对于识别目标来说是比较稳定的特征，因此我们的视觉信息处理过程中通过复杂的计算来追求这种稳定性，很大程度上补偿了光源光谱的变化，使我们在不同光源下观察物体的颜色所感到的偏差比实际的偏差小得多。我们对声音的相位很不敏感，但对于两耳之间的相位差极为敏感，这是由于相位差与声源方位之间有较为稳定的关系。经济性体现在特征抽取的正交性要求，即用较少的信息表现更大的范围，如果一种特征可以表现为其他特征的线性组合，那么这种特征是多余的。可预测性是认知主体从外界获得的信息品质的重要指标，特征抽取也不例外。对于动物

来说，构成重要事件先兆的特征无疑是重要的。例如我们在行走的过程中，由于道路的凸凹会使我们跌倒，而在跌倒之前的超越正常步行反馈所能应付的严重失去平衡是一个重要的特征，这个特征能使我们切换到应急控制系统以特殊的动作防止跌倒。当鸟类在飞行中要向树枝或地面等目标着陆时，需要在适当的时机将飞行动作改为减速着陆动作，这个最佳距离判断所需要的视觉特征就成为非常重要的特征，这个特征必须有助于准确地预测减速动作的结果。在特征抽取的算法之中，利用预测关系有时能取得很好的效果。从一幅图象之中将物体与周围景物分割开来，并从中获得物体的位置和形状等三维信息是非常困难的，需要极复杂的计算和大量的先验知识，而且结果很不可靠。但是在一个连续运动的摄象机上拍摄的多幅图象中利用预测的方法，从连续两幅图象预测第三幅图象，可以从各点预测参数的不连续变化得到准确的物体边缘，从预测参数得到正确的三维信息。大多数的脊椎动物眼睛长在身体两侧，两眼的共同视野很小，但是它们可以在自身运动的过程中对周围的三维世界作出正确判断，而且可以发现相对于周围静止景物运动的物体。对于一幅物体象按光的反射角度计算其三维信息通常可以得到两个解，如果我们把头固定住用一只眼睛盯住一个六面体方盒的一个角去看，并有意识地想象这是一个凹进去的角，就可以获得另一个解的感觉，但是只要头部稍作移动，这个幻觉立即就会崩溃，又恢复凸角感觉，这说明对于移动视觉的预测在确定三维关系中所起的重要作用。我们可以在几个人同时说话时只注意听其中一个人的话，这种特殊的“滤波”方式也可能与预测有关。

对于大多数动物来说，其行为方式与模式识别能力一样，大多数是先天的本能，而人的技能则多数要靠后天学习得到。象哺乳类和鸟类这些较高等的动物在具备 S 型学习能力的同时，也具有 R 型学习能力，即获得新的行为方式的能力，马戏团动物的精彩表演

就是 R 型学习的典型例子。粗略地讲,一种行为方式可以说是由多个单个动作组成的序列,或者叫作动作串,也可以看成一个程序,R 型学习就是通过学习生成动作串的过程,学习的前题是单个动作已经存在。把各种单个动作随机地组合成长列,然后判断其效果决定取舍,这种方法是指数爆炸的,因而不可取。学习动作串通常可行的方法是延伸与组合。延伸就是在原有动作串上加一个动作,使动作串一步步伸长。训练动物通常就是采用这种方法,用赏罚的方式把动作串中的动作一个个确定下来。但是由于种种理由,动作串的长度不可能也不应该太长,对于较长的动作串,只能采用分段的方法,一个个小串学习之后,再组合成大串。学习唱歌,乐曲演奏等都要采取这种分层次进行组合的方法。

完全指向外部的动作串通常不能使行为顺利进行,这种死硬的程序只能作为行为的框架,还需要加上反馈的修饰。在动物的行为中往往需要反馈修饰,在本能活动中反馈也是由本能提供的,例如大雁孵蛋时会用嘴把掉在巢外的蛋拨回巢里,这是个本能的动作,而在用嘴托住蛋时,为了防止蛋的滑落,显然需要精确的反馈,这种随附于本能动作的反馈也是由本能提供的。站立、行走和跑步都需要恰当的反馈保持身体的平衡,由于这些反馈的通路基本一致,从视觉和平衡感觉到下肢和全身肌肉,所以从站立到行走,从行走到跑步只需要调整反馈参数而不需要选择反馈特征和通道,学习起来比较容易。但是学习骑自行车时,由于反馈是从平衡器官到手,与走路时的反馈通道不同,需要一个试探反馈通道的过程,因而显得要困难得多。由于反馈要求较高的速度,不易由高层神经活动进行处理,只能通过低层神经模块完成,因此难以由意识进行控制,只能通过反复练习使反馈参数自动修正到最佳值,这是语言传授所无法代替的。反馈修饰有时也发生于动作串中各单个动作的连接处,以确定从一个动作转入下一个动作的最佳时机。

当动作串经过反复练习变为熟练行为时，反馈也会满足不了动作速度的要求，为了使行为更加连贯迅速，各个单个动作之间的关系变得更为紧密和相互渗透。在作一个动作时，已经同时在为下一个动作作准备，这种控制方式我们称之为前馈修饰或预测修饰。初学英文打字的人，是看一个字母打一个字母，熟练者可以以词甚至短句为单位，此时打一个词的动作已经形成熟练动作串，前一字母按下的手指放开的同时，手已经向下一个字母所需的位置移动，如果第三个字母在另一只手上，另一只手也已移到需要的位置。在我们说话时，由于这种前馈或预测的影响，后一个发音会使前一个发音产生变化，例如在汉语之中两个三声字连续发音时，前面的字会变成二声。在熟练和高速的动作串执行之中，反馈可以用来发现错误，却来不及立即修正错误。只能在其后的训练之中修正。

对于一个完善的动作串来说，还应该包括意外处理。意外处理是指在动作串的正常反馈机制无法处理的非正常情况下，中断原有动作串而转入的应急处理程序。这种中断出现在类似正在走路被石头绊一下，或是在骑自行车时有人在前方突然从左侧向右转弯这种情况。刚学会走路甚至稍大一些的儿童，走路跌倒的现象比成年人多，就是因为意外处理能力还比较差。由于意外处理的动作通常只有在意外发生时才能学到，而意外的发生频度又不高，所以意外处理的学习往往需要比较长的时间。意外处理的学习包括两个部分，一个是意外处理的动作或动作串本身，另一个是意外发生的检测信息。检测信息应尽可能提前给出，也就是说应具有预测的性质。以走路时的跌倒为例，跌倒的例外处理应在身体难以挽救地完全失衡之前加以检测并作出避免跌倒的反应动作，进而可以在脚踢到凸起物时还没有失衡就作反应，更进一步应在看到路面上的凹凸就作出相应反应。

由此可知，获得行为方式的 *R* 型学习中包含了复杂的多层次

的学习过程，首先建立动作串的框架，然后添加反馈修饰和前馈修饰，最后还要学习意外处理。这些过程之间又是相互渗透的，时时发生的意外处理会影响反馈修饰和前馈修饰，甚至修正动作串。例如刚学会骑自行车的人，有时会出现想要躲避障碍物或行人却照准障碍物撞上去的现象。这是学习过程的一个阶段，在学会控制平衡之后开始学习沿路骑行，此时人会紧张地盯着路面，所学到的反应是由意识控制眼睛盯住目标，而手作出相应的控制动作跟踪眼睛所注视的地方，其后随着熟练紧张程度降低，眼手随动不再起很大作用，但是当遇到障碍时又进入紧张状态，相似场合信息引出眼手随动的反馈，而此时眼睛却正好盯住怕撞上的障碍物，导致最终撞上障碍物。经过几次失败之后，这种由紧张引起的随动才会消失，使手眼随动的反馈在避障碍时不再起作用，只在过窄路时起作用。

在以上的分析中我们分别讨论了 *S* 型学习和 *R* 型学习，但这只是为了便于理解所作的区别，在实际的学习过程中 *S* 形学习与 *R* 型学习通常是密不可分的，*S* 中有 *R*，*R* 中有 *S*，*SR* 共存。我们学习的结果不是单纯的识别，也不是单纯的动作，而是感觉动作联合。在多数情况下，我们的感觉是一个主动的过程，这种主动性就体现在感觉所伴随的动作上。要得到触觉信息，我们需要用手去触摸物体，要得到视觉信息，需要把眼睛对准目标，要得到嗅觉信息，需要吸气的动作，化学家们还有用两个手指从试管口扇气味的动作，要得到味觉信息，需要把食物放入口中。表面看起来，听觉似乎不需要动作，人的耳朵不象猫那样经常转动，但是靠两耳的相位差从原理上只能判断左右方位，无法区分前后方位，要想分辨音源的前后方位，需要加上头部的转动，通过头部转动时相位差的变化，可以得到前后的信息。用立体声耳机听音乐时，由于头的转动不会引起相位差的变化，所以缺乏前后感觉，正前方的声音听起来象在头中间发出的。有一位机器人学教授曾

经在课堂上向学生提出一个问题：“树为什么不长眼睛？”这个问题初看有些古怪，但含意却是很深的，可以从很多不同的层次加以考虑。从比较浅的层次可以认为，树长眼睛也没用，假设一棵树看见有人拿着斧头和锯朝它走过来，它也没法逃走，只能眼睁睁地看着自己被砍倒。更深一层的问题是，树长了眼睛能否看见东西，有人作过这样一个实验，用一种装置检测人眼球的运动，然后用检测到的信号控制屏幕上的图形跟随被试验者眼球动作，使图形相对于被试验者的视网膜完全固定。经过一段不长的时间，被试验者感到图形融化得看不见了，但其他人却看见图形确实还显示在屏幕上。这个现象的原因可能是视神经的疲劳，它说明不动的眼睛是什么也看不见的。当然这种规则对于其他动物的眼睛未必适用，我们也可以假定树的眼睛在不动的情况下也能看见东西，或者树的眼睛是能象人眼一样运动的。即使如此，树也不会发现人朝它走过来，因为不动的树即使有了眼睛也不可能有距离的概念。动物的距离概念是靠运动与视觉信息变化之间的关系建立起来的，我们可以靠视觉判断距离是由于除了视觉之外我们还有运动能力，看到一个物体我们知道是否可以伸出手去抓住它，或是需要走几步之后再伸手去抓，距离概念只有通过这样的视觉加动作检验的过程才能建立起来，否则只靠两眼对准同一目标的过程只能得到两眼的角度差，而这个角度差到底代表什么意义，作为认知主体是无从知晓的，我们正是通过运动和由此造成的视觉变化之间的关系才得以建立距离概念的，只凭视觉，树可以看到一个物体越来越大，但不可能知道物体正在接近它。可以说，离开了运动能力，一个认知主体不可能得到有意义的感觉信息。

视觉的立体感比距离感更依赖于运动，通常用两眼看的立体照片，可以产生距离感，但很难产生立体感，这种照片看起来象舞台上的布景，成为前后位置不同的平面。这是由于我们平时的视觉立体感来源于头部位置变化所引起的物体投影形变，而看立

体照片时头部位置的变化不引起任何形变,所以立体感会消失。我们的立体感觉产生于视角变化(移动自身或旋转物体)与物体投影形变之间的关系,因此一个自身不运动的认知主体即使有了视觉也不可能具有立体概念或立体感觉。

关于感觉对动作的依存关系,有一个著名的动物实验——吊篮里的猫。实验装置是一根中间有垂直轴的横梁,可以绕轴在水平面自由转动,横梁的一头通过连杆拴在一只猫身上,这只猫可以沿圆形通道自由走动,象推磨的驴一样,我们称之为主动猫,横梁另一头吊着一个篮子,里面固定一只猫,不能自由走动,只能由主动猫带着转动,我们称之为被动猫。由于周围的背景对轴是中心对称的,所以被动猫所看到的景物和景物的移动情况与主动猫完全一样,也就是说两只猫所接受的视觉信息是相同的,所不同的仅仅是运动的主动程度。把两只幼年猫每天放在这个装置中数小时,那么几个星期之后,主动猫与被动猫显示出感觉和行动能力的明显差异。主动猫与正常猫一样可以选择落差较小的地方向下跳,而被动猫在不同落差面前只是随机跳下而不能作出正确的选择,这就是由于被动猫被剥夺了主动运动的机会,难以得到正确的距离感觉之故(Held and Hein, 1963)。

行为离不开感觉也是显而易见的。如果行为是由键刺激引起的,那么其起因就是感觉;在动作串中一个动作的完成引发下一个动作,动作完成的信息也来自于感觉;动作需要反馈修饰时,反馈信息也来自于感觉;前馈修饰是否达到了应有的效果,也要靠感觉来判断;意外的发生要由感觉来检测,意外处理的成败也要由感觉才能得知。双目失明的人行走比正常人困难得多,就是由于失去了视觉这个输入信息量最大的感觉,而其他感觉都无法提供足够的替代信息。

综上所述,感觉与动作是相互渗透、密不可分的整体,感觉中有动作,动作中有感觉,没有动作的感觉是未定义的感觉,无

感觉的动作是盲目的动作。我们所具有的知识是作为感觉动作连合的不可分离的整体，一个不可能引起任何内部或外部行为的键刺激是无意义的，“是什么”的知识归根结底要和“怎样作”的知识结合起来才有意义。这是分类的最基本原则，也是语义的根源。在计算机中，信息的意义要由人为的规定赋与，在生物的神经系统中，信息的含义取决于与之有关的输入和所引起的效果，也就是取决于该信息在动作感觉联合中所处的位置，而一种动作感觉联合的意义，则取决于它对于提高生物的适应度所起的作用。例如青蛙神经系统中由视觉处理得到的信息引起捕食动作，那么这个信息的含义就是由捕食行为给出的，而捕食又是一种“适应的”行为方式。象距离和方位这种基本概念则是由很多种感觉动作联合所共用的，对于适应起更大作用的信息。

4.5 智能模型与形象思维

用一个极端简化的模型可以把动作感觉联合表现为一个三元组 (S_1, R, S_2)。 S_1 代表初始状态下的感觉， R 代表动作， S_2 代表动作之后的终态感觉，即动作的效果。我们可以假定 S_1, S_2 和 R 连成一个总的模式存在自联想存储器之中，总模式中任一部分都可以从联想存储器中取出总的模式，即得到其他部分。根据输入和输出模式的不同，这个联想存储器可以有学习、动作、预测等不同的工作方式。学习工作方式的大致原理是，由一个随机信号源发出一个动作模式 R ， R 驱动执行机构（例如肌肉）对外界（例如关节角度）发生作用，使感觉（关节角度）由 S_1 变为 S_2 ，此时 R 和 S_1 经过一段延时等待 S_2 到达后一起被存入联想存储器，就完成了学习过程。这样经过多次学习的结果，联想存储器就具备了随意运动的能力。当上层模块需要完成某种动作时，只要对下层模块发出所要完成的目标感觉 S'_2 ，此后在下层模块已学

好的联想存储器中， S_2' 代替 S_2 与当时的感觉 S_1 一起作为输入模式，由联想得到 R ， R 输出到执行机构完成动作后，得到实际的感觉 S_2 ， S_2 与经过延时的 S_2' 进行比较，如果一致则表示动作完成，如果不一致则需要再学习。

以上所述的学习机制是学习的最低层的方式，也是最基础的方式，称为随机学习。其特点是所需要的先验知识较少，但学习效率较低，实际上动物的学习往往是以一定的本能为基础的，这样就可以大大缩小随机探索的维数和范围，而不象上述模型那样从“白板”开始。这个联想模型实际上也要求一定的先验知识，即哪些感觉应该与哪些动作产生关联。这样的先验知识是动物在进化过程中形成的一种合理的神经结构。例如可以用对老鼠训练踏板和电击之间的连合，即通过电击使老鼠不去踏某个踏板，也可以用制造呕吐感来训练老鼠不吃某种味道的东西，哪怕是在吃后 6 小时的催吐也可以通过训练造成联想，或称为条件反射。但是如果想把踏板与呕吐联合起来，或是在某种味道与其后的电击之间形成条件反射则非常困难，因为在老鼠的学习通道中不包含这种“不合情理”的联合。由于是低层次的学习方式，在成年人之中这种低效率的学习方式很不明显，大多数只用于小范围的参数搜索。但在婴儿身上我们经常可以看到一些随机的动作，这些动作是随机学习的表现。成年人只在万不得已的情况下才起用大规模的随机学习机制。

成年人随机学习的典型例子是在一个完全不会游泳的人落入水中时的学习过程。不会游泳的人落水时，由于呼吸困难的紧迫性，出现不容回避的困境，既没有可模仿的对象，也没有思考的余地，只能拼命乱动，以求通过随机学习在淹死前学会游泳。由于随机学习效率比较低，也可能有人学会游泳，但有的人还没学会就淹死了。对于这种落水后的胡乱动作，通常只被看成是“垂死挣扎”，不被认为是一个学习过程。但这确实是一个名符其实的

学习过程。五十年代我国一位有名的游泳教练穆成宽就曾采用过这方法教游泳，称为深水教学法。在要学游泳的人腰上拴根绳子，让他们跳入水中，教练拉着绳子的另一头，看看快不行了就拉一下绳子把学员提出水面再放下去，保护其不被淹死，这样可以使学员通过随机学习在很短的时间里学会将头部浮出水面的动作。

随机学习的一个重要结果，就是使人获得了随意运动的能力。由于我们的神经系统中的神经传导有上行和下行之分，上行的感觉信息是我们有可能意识到的，而下行的控制信息是我们意识所不可及的，我们可以随意地抬起手来，但是我们不知道抬起手时我们向哪些肌肉发出了什么样的神经脉冲，只能通过手的位置感觉得知手确实按我们的意愿抬起来了。在上面的 S_1RS_2 模型中，这意味着由上层来的目标 S'_2 ，汇同 S_1 一起得到控制信息 R ，从而执行了应有控制，得到与 S'_2 一致的 S_2 ，而 R 是我们感觉不到的下行控制信息。顾名思义，随意运动的本意是指由意识控制的或者说有意的、追随意识的动作。但是我们在这里所说的随意运动的含义要更广一些，既包括直接由意识控制的动作，又包括由意识间接控制的动作，也包括无意识的动作。这是由于此时由意识直接控制的动作，到彼时也可以间接控制，也可能完全脱离意识控制，这一点到讨论完意识之后可以更清楚。因此我们只考虑随意运动是由上层模块控制的动作，暂不强调与意识的关系。当在某一模块中通过学习形成了随意运动能力之后，上级模块对其进行的控制可以有两种方式，一种是给出 S'_2 引出 R 的方式，我们称之为目标驱动方式，此时我们可以假定在学好的模块中包含多种动作 R 。我们也可以假定一个模块中只含有一种动作 R ，此时可以认为含有多个 R 的模块是由多个单一动作模块集合而成的上级模块，那么上层模块对于单一动作模块的控制可以仅仅是一个起动信号，即命令这一模块从休息状态转入工作状态，此时不存在目标 S'_2 ，称为动作驱动方式。

动作驱动的随意运动可以进一步提高学习的能力，建立更丰富的感觉动作联合，例如有了原始的 S_1RS_2 之后，可以用动作驱动而非随机的方式学习到更多的在相同 R 条件下的新的 S_1 和 S_2 的联合。例如在幼儿行为中常见的抓握和拿近的动作，当幼儿通过某一个物体（例如奶瓶）学会了抓握动作之后，就可以用相同的动作去对付各种对象，从而获得新的感觉信息。他可以抓毯子、玩具等各种东西，在这个过程中建立视觉中的某种纹理和毛茸茸的手感之间的联合，建立视觉中的边缘和楞角的手感之间的联合，获得更为广泛的知识，使得视觉信息能产生“质感”。

动作感觉联合的另一个重要功能就是预测。在动作驱动的随意运动模块中存储了大量的 S_1RS_2 联合之后，在模块被启动时由当时的 S_1 和固定的 R 可以联想出 S_2 ，这个 S_2 是预测信息，把它与动作之后由感觉输入的实际的 S_2 进行对比，就可以得知预测是否正确，如不正确需重新学习加以修正，如果正确则说明模块有了较好的预测能力。当通过学习使得模块的预测能力充分发达之后，模块可以有一个新的工作状态，即对外界的输出 R 是可以关断的，而预测得到的 S_2 可以作为 S_1 重新输入模块作进一步的预测。此时我们并不需要对外界作实际的动作，只凭过去经验积累的感觉动作联合就可以对将要发生的感觉作出预测。这是感觉动作连合的纯预测工作状态，关断输出 R 也就导致了动作的内化，即在认知主体内部完成动作的效果。

维纳在《控制论》一书中谈到过雷达控制高射炮火时预测飞机轨迹的必要性，充分估计了预测在控制中的重要性。我自己对于预测的重要性的体会则来自一个很偶然的时机。在参加工作后不久时，我住在一个临时建筑的简易楼的二层，由于楼的一层在建成后重新加了一层水泥地面，使得楼梯的最下面一个台阶只有其他台阶一半的高度。由于当时灯泡紧缺楼梯又没有照明，于是这半级台阶引起了意外的麻烦，夜晚有客人来访时，对于常来的

人一定要提醒一下下楼时注意那半级台阶，对于初来的人则要打着电筒送到楼下。即使是老住户，不经意时踩到这半级台阶上也难免吃一惊。一个建筑物楼梯通常总是设计成各个台阶高度一样的，这样即使摸黑或不看着脚下下楼也不会感到太大的困难。下楼时试一两步就可以知道台阶的高度，以后按相同步态走下去不成问题，到最后一个台阶时再试一下是否到平地就行了。但是当最后一个台阶是半级台阶时，我们预测的平地或还有一级台阶都是错误的，当我们发现不是平地而按一级台阶高度向下踩时，提前到来的地面足以吓我们一跳。这个现象说明，我们不管作什么动作，都在时时刻刻预测着动作的结果，当实际的感觉与预测不符合时，引起我们的充分注意，或进入意外处理及触发学习机制。如果我们没有这种经常性的预测机制，“半级台阶现象”是难以解释的，只是当预测成功时我们通常感觉不到而已。

感觉动作联合是神经元模块的主要功能，因此和模块的组织一样具有分层和分级的组织结构，其全局学习过程也是逐层建构的。例如学唱一首歌曲，最低层的模块是直接控制声带和口腔动作的动作感觉联合，这里的感觉主要是关于音高和音色的。这些基本的动作感觉联合是在幼儿或少年期已经形成，是所有发音活动的基础，其结果是保证了发音的随意动作的可能，即我们想发什么音就可以发什么音。高一层的模块是一句歌曲的动作串，这一层的一个模块可以根据一句歌的前几个音预测下一个发音。更高一层的模块是关于整个曲子的，可以由一句的旋律预测下一句的旋律。由低层到高层的动作感觉联合，实现了由局部到整体、由表层到深层的预测能力。

这种逐层建构的预测能力，能够为我们提供用于思考的“智能模型”(Mental Model)，当我们能够对某一类事物的各种变化作出相对可靠而全面的预测时，我们就可以认为在我们的头脑中已经对该事物形成了“智能模型”，而智能模型的意义，也就是为我

们提供了多方面预测的可能性。Kenneth Craik (1943) 曾经提出过这样的想法：“如果生物体的头脑中持有外界现实的‘缩小的模型’，可以对模型试行各种操作，就可以使生物体具备多种能力。例如尝试各种各样的可能性，从中选择最佳的方案；在事件实际发生之前预想将来的事态而提前采取对策；把过去的事实中得到的知识用于现在或未来；在面临危机之际，从各方面采取更有效、更安全、更充分的对策，等等。”

与提出智能模型这一概念的 P. N. Johnson-Laird 一起工作的 Christopher Longue-Higgins 设计了一个简单的小车来说明智能模型的有效性。小车可以在桌面上来回移动，当它接近桌子边缘时就会响起铃声。但是这个小车上并没有装着任何一种直接检测桌边的传感器，小车完全依靠模型来探知桌子的边界。由小车主轮带动两个小轮，小轮带动一个与桌子形状相同的小纸片运动，小轮与小纸片的相对关系总保持着和小车在桌面上的位置相互一致，这样当小车到达桌边时，小轮正好触及纸片边缘，从而接通电铃。Johnson-Laird 认为，认知科学家中几乎没有人会怀疑内部模型的有效性，问题在于，这样的模型应具有什么样的内部表现，以及怎样在认知过程中使用这些模型。我们认为这个看法很对，但是漏掉了一个重要的方面，即这些模型是怎样被制造出来的。我们关于动作感觉联合及预测的讨论，可以说已经对上述这些问题给出了初步的解答，包括了模型的表现——动作感觉联合，以及模型的生成和修正所涉及的学习方法，也包括模型使用方面的一些内容。由于 Johnson-Laird 忽视了模型制造的问题，所以他的小车只能在一个桌子上走，到了另一个桌子上原有的模型就失效了，必须由人再作一个纸片。如果按我们的观点重新设计小车的话，应该考虑模型的生成与修改的问题。完全无感觉的认知主体不可能获得任何知识，设计能够自己获得客观世界模型的小车应使其具有最低限度的感觉，例如在前方和左右各有一个

触角，可以在到达桌边时感觉到，小车可以从内部模型是白纸的状态下开始行动，当触及桌边时就在同步运动的白纸上打一个黑点，再转一个方向行走，直到得到桌面形状的完整模型，使用这个模型时用光电器件读纸片，遇到黑点就是到达了边缘，如果在纸片上对应小车行进方向上装一个超前的光电器件，那么在小车行走时在到达桌边之前就可以预测出前方的桌边。有了这种预测能力，小车的行动速度可以大大加快，因为在到达桌边之前可以提前减速，不至于由于刹车不及而掉下去，而只依靠触觉的小车就只能缓缓而行，摸索前进。以上的学习方法是随机的，效率很低，要得到由点构成的桌边线需要长时间的学习。效率更高的方法是，遇到桌边时转一个直角，用一侧的触角作为引导沿桌边走一周，可以更快的制成内部模型。这个方法假定桌子中间是没有洞的，但是只要预测的失败可以触发新的学习过程，那么不管桌子中间有几个洞，最终是可以得到正确模型的，而且在把小车放到另一个桌子上时，同样的原理也可以触发新的学习过程。为了使小车不过于保守，小车应设计得可以偶犯错误而越界行走，这样在小车放到更大的桌面上时，不至于永远留在小圈子里打转，可以有机会获得更大的自由，这是 *ES* 学习的基本原则之一。

我们用一个实际的例子来说明使用智能模型进行思考与通常所谓的逻辑思维的区别。我们考虑这样一个问题：有三个人 *A*、*B*、*C*，*A* 在 *B* 的左边，*B* 在 *C* 的左边，那么 *A* 是否在 *C* 的左边？这个问题如果用逻辑来解决的话，首先要考虑“在左边”的关系在逻辑上是否可以递推，就象 $A > B$ ， $B > C$ 则 $A > C$ 那样。如果可以递推，则 *A* 在 *C* 的左边，如果不能递推，则从逻辑上是得不到解答的，这个关系是否能递推是不明确的。利用智能模型的方法，可以在头脑中设想各种实际的空间分布，来考虑这个问题的解答。例如可以认为三个人是围圆桌而坐，如图 4.6。从不同的图中很容易判断 *A* 和 *C* 的关系，在图 4.6a 中 *A* 在 *C* 的左边，在图 b 中，*A*

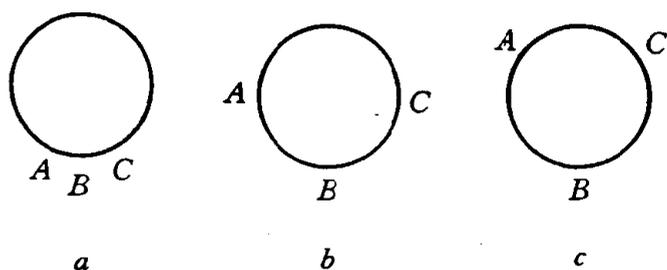


图 4.6

在 C 的对面, 在图 c 中 A 在 C 的右边。通过这样的形象化的分析, 我们甚至可以为“在左边”的可递推性 (满足推移律) 找到一个充分条件, 即在递推过程中, 每个人正面方向的夹角之和小于 π , 而且每个人的位置偏离正左方向角度不大于正面方向夹角的一半。

从这个例子可以看到, 用逻辑的方法无法解决的问题, 用智能模型是可以解决的, 而且可以为逻辑提供基础和引导。“跳蚤是动物, 所以大跳蚤是大动物。”这个推理我们一看就知道是错误的, 虽然我们未必知道这个推理违反了哪条逻辑推理规则, 或是使用了哪条不正确的推理规则。我们发现这个结论的错误, 是靠了一种形象的冲突, 大跳蚤可以想象有米粒或豆粒那么大, 但大动物却使我们联想起象、鲸、甚至恐龙。这个判断是靠智能模型完成的, 我们通常所说的形象思维, 就是靠智能模型进行的思维活动。但是形象思维这个说法比较侧重于视觉形象, 而智能模型则更为普遍, 作为动作感觉联合, 可以包括多种不同感觉的综合效果。从这个意义上讲, 形象思维应该推广成模式思维才更为全面, 当然, 由于视觉是各种感觉之中信息量最大的一种, 所以在模式联想思维之中, 形象思维占了主要的部分, 我们也可以用形象思维作为模式联想思维的代表, 这样以后我们谈到形象思维时, 只要注意它不仅仅局限于视觉领域即可。

形象思维是对智能模型的操作，是内化了的动作，是利用感觉动作联合进行的一连串的预测。关于形象思维的存在性证明，有一个经常被引用的著名心理学实验。实验是向被试者出示如图4.7那样的成对的立体图，要求被试者判断两个图所表示的立体物是否是同一物体。实验结果表明，判断所用的时间与两图之间相差的角度呈明显的正比关系。从而说明人在进行这种判断时所采用的方法是在头脑中以一定速度旋转其中一个立体物，然后判断是否能够通过旋转使一个立体物与另一个重合，以此证明人的头脑中

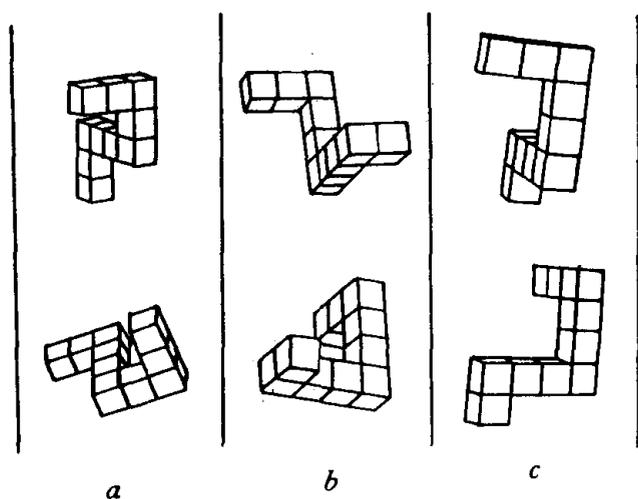


图 4.7

有对形象进行操作的能力。图4.7中 *a* 和 *b* 是相同的，而 *c* 是不相同的。这个实验对于形象思维的存在是很有说服力的，但迄今为止对于形象思维的机制方面说明得还很不充分。

我们可以使用这个例子对形象思维的机制作一些更为具体的讨论。这个实验的巧妙之处在于，所用的立体之间没有明显到可以一眼看出的特征差异，立体之间的差异只是左旋和右旋的区别。而左旋与右旋的差异又是非通过旋转重合难以判断的，如同我们在电磁学中所用的右手定则，往往还真需要伸出手来比划一下才

行。需要说明的重要事实是，这种心理旋转所需要的时间为什么与要旋转的角度成正比。这个实验结果意味着，心理旋转是逐步进行的，每一次只旋转一个小角度，而不能一下子转一个大角度。要解释这种现象，需要先弄清立体感的内部表现。前面我们已经把立体感看成是一种感觉动作连合，即我们自身视点移动所引起的视角变化与物体二维形象变化之间的联想关系。也就是说，理解一个立体，意味着能预测这一立体物在自身旋转时造成的不同的平面象（旋转轴与视线不平行）。为了实现这种联想，我们可以想象大脑皮层的视觉处理部分有一些专门处理运动视觉的模块，每个模块对应视网膜上的一个点，通过一定的时间和周围近处模块之间的相关关系得到局部平移信息，这些局部平移信息再与更大范围的模块相联系，进而得到平移速度差，再进一步由高层模块得到整体的旋转信息。由于感觉动作联合具有预测能力，旋转动作驱动也可以引起各局部模块预测出旋转所引起的二维图象变化。由于立体感需要由不大的旋转角度得到，以提高立体视的灵敏度，因此利用这一感觉动作联合所作出的预测也将是小角度转动的预测。因此大角度心理旋转需要由多个小角度旋转来完成，要费较多的时间。

以上这个实验如果对同一个人反复用几幅图作实验的话，久而久之，被试者会记住图的形状和结论，从而可以不经旋转，直接给出判断结果。用一个更明显的例子来说，准备五六个立体，各自涂成不同颜色给被试者看，一开始被试者要经过心理旋转才能判断两个物体是否相同，但几次判断之后，在被试者头脑中会形成新的感觉动作联合，认准了红的和绿的一样，绿的和兰的不同，就可以直接作出判断而无需心理旋转了。这种现象我们称之为“短路效应”。正是由于有了短路效应，才使得人的思考是知识越多，思考越快，而不象传统的人工智能系统那样，知识越多，思考越慢。思考所得到的结论可以被记忆，而下次对同样或同类的

问题可以直接得到结论而无需重复思考，这是能学习的机器胜过死板机器的优越之处。

大多数的动物依靠本能而生存，即使有一定的学习能力，也是靠本能提供一定的学习算法和预备知识。人的学习能力比其他动物强得多，人的生存技能的很大一部分要靠学习才能得到，学习的方法也比其他动物丰富得多，而且前期的学习可以为后期的学习提供预备知识，通过学习也可以改变和获得学习算法。因此，具体地研究人在每一个学习过程中的学习算法既复杂又可能是缺乏普遍意义的。但是我们仍然需要而且可以找到一些具有普遍意义的基础性的学习方式，如果没有最基本的由本能提供的学习能力，更具体的学习算法是无法获得的。为了讨论人的复杂的学习过程，我们把人的学习分成三个层次，即随机学习、模仿学习和思考学习。

随机学习在前面已经讨论过，它是学习中最低层次的一种方式。随机学习的特点是要求的预备知识很少，几乎可以从无到有，但是学习效率较低，需要限制学习范围以避免指数爆炸。学习的结果是使认知主体建立低层次的直接与外界环境相关的感觉动作联合。感觉动作联合的建立使认知主体获得随意运动的能力和简单的预测能力。随意运动的能力又可以大大提高认知主体对于客观环境的探索效率，使得感觉动作联合更快地丰富起来，逐步聚合成较为完善的智能模型，大大提高预测能力。从这个过程可以看出，动作感觉联合是具有自我增殖能力的，这种增殖能力依靠的是一种泛化机制，或者叫作无限推广倾向，抓奶瓶的动作同样可以试着用来抓毯子。泛化是一种普遍的机制，即一种方法可以对付不同的对象，在更高层次的学习之中泛化机制也是很重要的。随机学习不仅具有自我增殖能力，而且由学习所得到的随意运动能力和预测能力为更高层次的学习准备了必要的条件。虽然随机学习从原理上讲也可以实现逐层建构，但是由于其效率较低，在

更高层次上模仿和思考的学习方式显得更为优越。当然，在更高层的学习中仍然也包含随机因素，正如随机学习也受着本能约束一样。

模仿学习是人类学习过程中最常用的学习方法，我们的教育体系之中大多数的教育活动是通过模仿学习的机制实现的。模仿必需以随意运动能力作为基础，也就是说，看到别人的某种行为时我们要具有能够实现相同行为的能力，否则模仿当然是无法进行的。模仿学习的另一个必要条件是模仿的欲望，通常在灵长目动物的行为中已经可以发现明显的模仿能力，某些鸟类的发音模仿能力也相当强，但人类发育过程中的模仿能力在范围和深度方面都远远超过其他动物。模仿学习的特点是其学习效率高于其他任何一种学习方式，这是由于在模仿过程中随机探索的因素较少，由于存在着模仿的范例，即使出现偏差，其修正方向也是确定的。当然在模仿学习中也不可能完全避免随机探索的因素，因为认知主体对范例的观察不可能是完全的，例如我们可以看到别人完成体操动作的姿势，却观察不到作这种姿势所用的力量。很多只可意会不可言传的东西仍然需要探索，学习骑自行车时的平衡反馈就是一例，反馈参数是既不能靠观察得到，也难以用语言表达的。

根据短桥原理，对于复杂的技能，模仿学习也需要通过逐层建构的过程，不可能一蹴而就。唱歌需要一句一句学，高度的专门技能有时需要几年、十几年才能掌握，这样一些复杂的技能需要由多次小规模的学习积累而成，局部的模仿学习结果为全局的模仿准备条件，低层次模仿的成功为高层次模仿打下基础。模仿这种学习方式往往给人以创造性不足的印象，但是由于模仿的高效率，可以使人在短时间里获得丰富的动作串，由此又能建立大量的智能模型，使我们不必事事都要经历困难的发明过程，而能够在相对较短的时间里积累庞大的知识。这个庞大的知识结构是我们进行创造性思考的不可缺少的先决条件。模仿学习的最重要

的意义在于,模仿能力为文化的传播和继承准备了必要的条件,对于没有模仿能力的动物来说,文化现象是不可能产生的。文化可以使技能的发展突破基因进化的世代时间约束,从而得到比本能技能高得多的发展速度,因此生物越是高度进化,硬基因增加,发育期和寿命增长,本能的进化越慢,文化的利益也就越大。在一些高等动物之中,已经可以看到文化的萌芽,但是人类的文化体系却远远超过其他任何动物。人类个体的智力与文化体系组成软硬结构,其中智力是软结构,文化是硬结构,人在文化环境的支持下通过模仿获得高度的智能,而又通过创造性思考对文化环境作出建构,可以说,对文化的模仿和创造是人类文化高度发达的主要原因。在此之前,人类从猿到人的进化过程中,文化与智力之间的相互作用可能是智力高度进化的主要原因。某种水平的文化财富使得模仿能力较强的类人猿个体得到较高的进化适应度,而这些类人猿又能在较高的水平上创造更高度的文化,这种智力与文化之间的互适应协同进化,可能是从类人猿向智人进化的主要动力。

比起模仿学习,思考学习具有更大的创造性,而思考学习的效率又高于随机学习。思考学习所需要的前题条件是高度的预测能力,而这些能力来自于丰富的感觉动作联合和大量的智能模型。因此思考学习要求随机学习和模仿学习的结果作为其基础。思考可以利用预测能力组织新的动作串,而无需外部的实际动作或范例。例如我们需要一个动作串去完成从 S_1 到目标 S_g 的变化,此时可以探索 S_1 到 S_2 的各种预测,如果对于某个 S_2 存在着 $S_2R_2S_g$ 的感觉动作联合,那么把 $S_1R_1S_2$ 和 $S_2R_2S_g$ 联成一个动作串 $S_1R_1R_2S_g$,则 R_1R_2 就是我们需要的动作串。思考过程是靠操作智能模型进行预测完成的,这种操作是已经内化的动作,不引起实际的外部动作,因此其能量效率和时间效率都比较高。当我们把一个动作串看成一个高层次的动作,即 $R=R_1R_2$,那么通过思考

我们就得到了新的高层次的感觉动作联合 S_1RS_g ，这种短路效应使得 S_1RS_g 成为今后的思考可利用的感觉动作联合，更可以通过泛化的手段由这个感觉动作联合推广成新的智能模型，使今后的思考效率更高。因此思考的过程也是可以逐层建构的广义进化过程。上述的创造过程从一次思考来看也是典型的指数爆炸的搜索，但是只要遵循短桥原理抑制搜索步数，就可以通过广义进化的多次积累达到大的创造。

在人的实际的学习过程之中，随机学习、模仿学习和思考学习这三种学习方式并不是逐一单独进行的，而是你中有我，我中有你的关系。随机因素在初期建立低层次感觉动作联合时比较明显，随着模仿学习成份增加而减少，但在思考学习之中又有所增加。模仿学习即使在思考能力比较强时仍然是重要的学习方式，因为创造性思考只有在和模仿一起与文化环境发生相互作用，在软硬结构之中才能取得最大的效果。每一个具体的学习过程，分析起来都可能包含这三种学习方式，只是在学习的不同阶段各自所占的比例不同而已。在模仿的过程中往往包含思考和随机搜索过程；而思考中的目标设定往往与模仿有关，思考的过程也基本上采用随机搜索的方法；因此对于学习过程要从这三种方式去分析才能得到比较全面的印象。

逐层建构的原理可以使我們回避指数爆炸所带来的困难，但也由此而产生了这样一个问题，我们的大脑之中模块组织的层次显然是有限的，那么我们对于外部世界所建立的模型的深度是否也是有限的呢？这种层次深度的限制在动物的学习之中是经常可以见到的，对于有学习能力的动物，我们也难以通过训练使其无限地聪明起来，动物的学习算法通常是 L^1 型的，通过学习改变学习算法的可能性不大。而人的学习能力却显示出明显超越其他动物的特征，看起来象是 L^∞ 的学习算法，这种 L^∞ 的学习又是怎样实现的呢？或者说，我们怎样在有限的大脑中实现无限的认知深

度呢？这种现象与人的外部表现能力有关。与其他动物相比，人具有强得多的外部表现能力。虽然很多动物都具有一定的相互通讯能力，但是这些通讯内容往往是只具有固定的意义，在进化中产生的特定通讯通常是直接地受到适应度的约束，其自由度是很小的。例如当某种昆虫发出特定的气味或声音来吸引异性时，发出信息的差错和接收信息的差错都会使昆虫的适应度下降，进化使得这些有适应意义的信息传递方式高度硬化和特化，失去了表现丰富内容的可能性。在人的进化过程之中，在智力与文化的互适应协同进化产生之后，在文化支持智力的模仿能力逐步发达的同时，创造力和表现力这种智力对文化的建构能力也成为与适应度密切相关的因素，获得了相应的进化压力，使得人类在进化过程中得以逐步提高其表现能力和表现欲望。这种外部表现能力与其他动物的通讯能力的主要区别就是有很高的表现自由度，离开这种自由度，文化建构是不可能的，因为文化现象本身的意义就在于它具有基因所不可决定的内容，基因除了提高通讯自由度之外，没有其他支持文化的方法。在远古时代人类居住的洞穴之中就已经出现了壁画，这是早期人类表现欲望和表现能力的证明；语言和文字的出现使表现的效率得到飞跃的提高；卡拉OK的流行也是由于它为满足人的表现欲提供了一种简易的手段。

如果把思考看成是动作的内化，则表现是感觉的外化，是认知主体之中下行的信息传递。这种下行的外化与上行内化所形成的环，就提供了无限思考深度的基础，使得思考的深度不依存于模块组织的层次深度。表现能力从一开始是在认知主体与文化环境的相互作用中发展起来的，在与其他人对话过程中掌握语言，在与实物的对照之中学会绘画，在与他人对弈的过程中学会下围棋，在与他人争论的过程中学会辩解。但是表现并不仅仅是向着外界的，我们可以听到自己的语音，看到自己的绘画，看到自己下一步棋之后棋盘上局势的变化，评价自己辩解的理由是否充分。随

着表现能力的发展，表现的反馈能力也发展起来，在认知主体内部可以形成文化环境的智能模型。我们可以并不发出声音而听到自己在内心里说些什么，可以不用动笔闭着眼睛想象出一幅图画，甚至可以不用棋盘下棋，可以自己和自己争论。无论多么深层的思想，只要可以通过语言或其他形式加以表现，那么就可以通过表现反馈通道出现在浅层感觉领域之中，进一步被加工、被改造。因此，模块组织层次深度的有限性，并不造成思想深度的有限性，这是 L^∞ 型学习之所以成为可能的基本根据。

在人类的各种外部表现形式之中，语言由于其效率高、自由度高、表现能力强而具有特别重要的地位，图形作为外部表现比语言效率低得多，但由于对于某些对象来说有语言所不可替代的作用，也是一种重要的表现形式。在思考过程中作为浅层感觉的内部表现，语言与形象的效率是相差不大的，在内部视觉中生成一个形象并不比生成一句话慢，因此在思考过程中使用形象还是语言则取决于思考的对象的性质，以及与思考对象有关的智能模型涉及哪些感觉。由于我们思考所使用的智能模型有些涉及视觉形象，有些涉及语言，因此有不少人认为思维有两种形式，即形象思维与逻辑思维，形象思维的所用的内部表现是形象，基本操作是形象的联想变换，而逻辑思维则所用的内部表现是符号，基本操作是逻辑推理或按照语法规则进行符号变换。这种分类法所导致的困难在于，形象思维与逻辑思维之间在理论上很难建立起有效的相互关系，而且如果实际观察思考过程，我们利用语言作为介质的思考，虽然具有其合理性，但是这种合理性类似于在进化中产生的生物结构的合理性，它是合理的，但未必合乎形式逻辑。在我们现实生活的对话之中，虽然明显不合逻辑的并不多，但是要想在前后的语句中找出严格合乎逻辑的形式化推论过程，也几乎是不可能的。反过来，用逻辑来处理或解释语言现象，也是不可能成功的，因为形式化的逻辑只能涉及语法，而不可能涉及

语义，而语义在语言现象中当然起着举足轻重的作用，即使仅就语法而言，其中合乎逻辑的成分也不是很多的，在自然语言中，可以说没有一条语法是没有例外的。因此把以语言为介质的思考活动称为逻辑思维是一个概念错误，如果认为语言是逻辑推理的产物，就无法解释人为什么会说出不合乎逻辑或者并不严格合乎逻辑的话来。因此在我们对认知问题所分的层次之中，逻辑与数学、物理学一样，是建筑在语言基础之上的文化产物，没有语言就没有逻辑，但逻辑既不是语言的基础，也不是思维的深层规则。所谓“逻辑思维”仅仅是一个错觉。

排除了“逻辑思维”在思考中的原理性解释之后，我们需要从形象思维的角度去说明语言在思维过程中的重要作用。如果把形象思维看成是模式联想思维的简称，那么借助于语言的思维也是形象思维的一种。在这里我们不把语言看成是符号的集合，而看成是模式，因为语言不管表现为听觉的发音串还是视觉的文字串，都是具有空间和时间分布的信息。因此从基础的原理上看，借助于语言的思考与借助于视觉形象的思想是可以统一在模式联想或感觉动作联合之中的。但是语言又有与其他感觉模式的不同之处，就是它的间接性。除了拟声词与实际的感觉直接相似之外，大多数的语言模式与语义之间是没有必然内在联系的，因而语义就成为语言得以成立的一个重要问题。对于动物来说，键刺激的意义是由这个键刺激所引起的动作以及该动作对于动物生存的意义所决定的。然而语言在很多场合并不引起直接的外部动作，但是语言可以影响人的思考过程，而思考则是内化了的动作。因此语言的意义也就取决于它在认知主体之中能够或实际上与哪些智能模型相关，引起对智能模型的哪些操作，能够联想起哪些感觉动作联合。

对于生活在同一语言环境、文化环境和现实环境中的人来说，由于语言与智能模型之间的联系具有一定的共通性，使得人们很

容易通过语言达到沟通和理解，语义也就具有一定的稳定性。但是同一个词汇并不一定在任何情况下都只与一个智能模型有关，因此就造成了词汇的多义性。在实际的语言运用过程中，由于语境的约束，这种一词多义的现象并不会造成语义沟通的障碍。语境可以看成是一种稳定性比较差的临时建立的智能模型，永久性的智能模型在一开始也是由临时性智能模型在反复运用的过程中强化了多数情况下有效的部分、淘汰掉不稳定成分逐步建成的。由于语境的临时性和不稳定性，具有较大的个人和集团差异，造成语言理解的困难。这种困难有些是起因于不具备理解该语句的智能模型，例如外行听到某一专业的行话时会无法理解；有些则起因于缺乏必要的语境约束，有关的智能模型过多而无充分信息帮助选择，例如以下两个人的对话。A：“那事怎么样了？”B：“杨明那边已经通了。”对于参加对话的A，B两个人来说，由于语境信息充分，相互可以充分理解，但是对于旁听的第三者，则只可能引起大量的猜测而无法确定其语义。“那事”可能是A委托B办的一件事，也可能是A，B曾经同时参加的一项工作，也可能是B向A讲述过的一件没有结局的事情等等；“已经通了”可能是指一个命名为杨明的隧道已经打通，也可能是叫杨明去打通关系已经成功，也可能是杨明正在编的软件已经调试通过，有很多种可能性而无法确定。

对语言的理解意味着听到语句之后在认知主体内部能够联想起相应的智能模型的操作，而说话又是智能模型逐层控制产生的动作，当这种动作关断了最下层的对肌肉输出时，只剩下了预测性反馈的作用，使我们自己能听到自己没说出口的话，并且对这些话进行理解和评价，据此可以修正和再建智能模型，通过思考提高思考的能力，这就是以语言为介质的思考过程。由于我们只可能意识到上行的感觉信息，意识不到下行的控制信息，所以对于动作只能依靠感觉的差异进行修正。出于同样的理由，对于思

考这种内化了的动作，也只能依靠操作智能模型的结果、即生成的语言和形象来修正或建立智能模型。在语言现象之中，比喻是一种很常用的手法。比喻从字面上讲就是通过类比的方法使人理解，但是比喻的机理用逻辑的方法却是难以理解的。从语言的理解与智能模型的关系出发，可以对比喻的作用作出解释。假定 A 要向 B 表达一个感觉动作联合 S_1RS_2 ，由于 R 是 A 自己也无法意识到的动作，只能向 B 表达 S_1 和 S_2 ，而由于 B 的头脑中不存在相同的动作感觉联合 S_1RS_2 ，因而对 S_1 和 S_2 感到难以理解。但是在 B 的头脑中并非不存在 R ，只是 R 是以 $S'_1RS'_2$ 的方式存在着，用 S_1 和 S_2 形成不了对 R 的联想。由于 A 具有比较完全的智能模型，可以向 B 传递这样的信息： S_1 与 S_2 的关系和 S'_1 与 S'_2 的关系相同，由于比喻信息 S'_1 和 S'_2 的存在，与 B 的头脑中的感觉动作联合 $S'_1RS'_2$ 发生联想，得以引出 R ，并与 S_1 和 S_2 联合，产生新的动作感觉联合 S_1RS_2 ，使得 S_1S_2 成为可以理解的信息。这就是比喻有助于理解的作用原理。

把形象思维看成是对智能模型的操作这一观点，有助于解释关于记忆的一些困难问题。诺尔曼在关于记忆的讨论中提出过这样的问题：“在不知道必要信息是什么的情况下，通过检索发现这个信息时，怎么会知道它就是要检索的信息呢？例如，当我们努力回忆起多年以前去世的同事的名字时，突然想起 Isac Newton，在没有想出同事的名字时，又怎么知道他不是 Isac Newton 呢？这个例子包含了解决问题的关键原理。”这种现象在生活中是很常见的，我们可能费很大劲也想不起某个中学同学的姓名，但是如果别人说出一个姓名，我们却可以判断是不是我们要想起的那一个。从简单的逻辑来看，这种现象不合乎逻辑，如果我们知道某个姓名，当然很容易通过比较得知和这个姓名不同的姓名都不是这个姓名，而当我们不知道正确答案时，我们就无从比较和判断。如果把姓名看成孤立的符号，上述问题确实是无法解决的，但是如

果把姓名看成是一个智能模型中的一个信息，情况就不同了。我们可以认为关于某一个人的各种信息构成一个智能模型，如果这个智能模型比较完整，我们可以对这个模型进行操作，从姓名可以联想起其音容笑貌、他的服装、性格、趣事、外号等各种信息，也可以从其他信息联想起他的姓名。当某个人停止与我们接触之后，随着岁月的流逝，关于他的智能模型也被磨损得残缺不全（这个过程与学习的机制有关）。这种不完整的智能模型可能会出现这种情况，即我们可以由姓名联想起他的面容，却不能从面容联想起姓名。当我们极力回想某人的姓名时恰恰就是这种情况，即他的面容及一些其他信息为我们所知，但从这些信息已不能直接联想起他的姓名了。但是从姓名到其他信息的联想却不容易被磨损，这是因为面容是我们进行形象思维区分个人所用的主要模式信息，而姓名在思维过程中主要用来引发面容等其他信息，从面容联想起姓名的过程使用频度要小得多，因而更快地被忘却。在这样的不完整的智能模型之中，我们很容易判断一个姓名是否是我们想要回忆的那个人，因为如果是那个人，那么这个姓名可以联想起已知的面容，乃至更多的属于该智能模型的信息，而不是这个人的姓名则没有这种效果。这就象对于一个难解的方程，我们不知道它的根是什么，但是如果给我们一个数，可以很容易地把它代入方程来判断它是不是这个方程的根。

4.6 感情与评价

感情与智能的关系和感情在认知过程中的作用长期以来是科学研究的空白，没有受到应有的重视。由于逻辑主义和符号主义的影响，人们往往把认知和理性的核心问题看成是真和伪的问题，在判断真伪的过程中，感情处于无足轻重的地位，甚至是一个干扰和妨碍的因素。文学家们也很热衷于描写感情与理性的冲突。但

是如果我们把认知的核心问题看成是认知主体建立内部的智能模型的优劣问题,是一个评价问题,那么感情作为价值观体系就具有举足轻重的地位。从根本上讲,崇尚理性本身也是一种价值观,所谓感情与理性的冲突,实际上只不过是生物价值观与文化价值观之间的冲突的表现而已。从这个观点出发,我们可以认为感情是认知、智能和理性的领航员,起着决定方向的作用,因而在与认知有关的研究之中,感情是一个不容忽视的起着关键作用的因素。

对于不需要学习或只进行固定学习算法控制的学习的动物来说,感情并没有存在的必要。对于这些动物来说,行为方式或一定程度的学习方式是由基因决定的本能实现的。这些动物行为方式的改进由进化过程完成。由于进化过程本身具有明确而单纯的评价原则,即达尔文的适应度,所以这些动物并不需要再具备其他的价值体系。但是对于学习能力较强的动物,特别是像人类这样学习所得到的技能远远超过本能,学习内容丰富而复杂,学习算法也可以由学习产生的认识主体来说,感情作为评价体系就成为必不可少的先决条件了。这是因为进化与学习虽然都遵循变异与选择这一相同的原理,但是进化中的选择可以通过不适者的淘汰即死亡或不育来进行,而死亡对于学习来说不起任何积极作用,因为学习是生存期间中的过程,动物不可能把死亡作为学习中的选择基准,因而学习的选择就要求有在生存期间中起作用的评价体系,这个评价体系就是感情。因而感情现象与学习能力在动物中呈正相关关系也就决非偶然了,人是学习能力最强的动物,同时也是感情最丰富的动物。

评价的高低(正负值)就表现为感情上的好恶和行为上的趋避。一个价值体系的简化模型可以表现为一个价值向量,向量中的每一个元素可以代表一种价值,元素可以取正值和负值,正值代表正的评价,可以引起喜好的感情和趋向的行为,负值则代表负的评价,可以引起感情上的厌恶和逃避行为。评价值的绝对值

越大则表示对生物有适应上的意义或认知上的意义，绝对值为零的评价则代表无意义的因素，不引起生物或认知主体的注意。如果用这种简单的模型来表达生物头脑中或认知主体中的感情运行机制，可以认为生物或认知主体中由一个评价向量表现其当时的整体感情状态，而每一个模式信息中都带有一定的感情色彩，可以是与评价向量对应的一个子向量。模式信息中感情子向量与整体感情状态向量之间的相互作用和相互影响则构成了感情状态向量的变化，以及通过学习改变模式信息中感情色彩子向量的根据和机制。

我们的感情状态向量到底有多少个元素，或是说我们人类一共有多少种感情，是一个尚未得到充分研究的问题。中国古代有七情六欲之说，现代的情绪心理学也有各种不同的说法，对此可以暂时不下结论，至少当我们把价值观也作为一种文化现象来考虑时，有理由认为对于不同的个人来说感情状态向量的维数有可能是不同的。人的感情评价系统按其来源的不同可以分为生物评价系统和文化评价系统两大部分。生物评价系统来源于人类的进化过程，在个体中被基因所决定，表现为人的天生的性格特征。每个人的性格的区别以多样化的价值为根据，而人类生物价值观的共性则由共同的进化过程所保证。生物价值评价系统和一切其他性状一样，在进化过程中由随机因素而产生，由其对适应度的贡献而被维持。对于不进行学习的动物来说，这些适应的性状表现为某种功能结构或固定于基因的行为方式，而对于善于学习的动物来说，由于其行为方式不固定于基因，而在后天学习而得到，因此适应性状只能在基因中表现为作为学习导向的生物价值观，即趋利避害的方向性指引。由此可知生物价值观在进化过程中的形成机制与其他性状是完全相同的。例如在鹰鸽博弈模型之中，对于不学习的动物来说，鹰战略直接体现为引发具体的战斗动作，鸽战略直接引发具体的逃避或臣服动作，这些动物的行为方式是固

化在基因之中的；而对于靠学习获得行为方式的动物来说，鹰战略不是直接引起战斗动作，而是引起战斗欲望，引起一种斗志，一种激愤情绪，而具体的战斗动作则是要在战斗中逐步学习的，为了引导这种学习，则需要重创对手时的快感，自身负伤时的疼痛感等一系列具有感情评价色彩的感觉。在经过学习获得了一种或几种战斗动作时，由于这些动作都是在战斗情绪之下学到的，具备与战斗情绪之间的模式联想（情绪也是一种模式），战斗情绪就可以引发这些战斗动作。这是心理学中熟知的一种现象，即某种情绪可以引出与该种情绪有关的联想，悲伤时容易想起过去悲伤的事件，欢乐时也容易想起欢乐的事件。

文化评价系统的产生需要一个当然的前题，即文化现象的存在，这意味着该种动物具有足以支持文化现象存在的智力本能。这种本能主要体现为对于文化的模仿和重建的能力与欲望，同时也体现在文化价值观的传递能力上。文化进化的基本原理与生物进化是相同的，都是依靠变异和选择进行的，但是又有其各自的特点。现代进化论的发展逐步否定生物进化中集团选择的因素而重视个体选择甚至基因选择，走向“自私的基因”这个基本原理，但是由于文化本身是超个体的，主要是集团的现象，在文化进化中个体选择虽起一定的作用，但集团选择往往占有主要地位，当一个文化集团在经济或军事上处于强盛地位时，它的文化价值观体系及意识形态和艺术风格都会具有较大的渗透能力。当然在文化进化之中个体选择也还起着重要作用，例如一种新产品能否在市场上站住脚基本上取决于用户的个体选择。文化进化的另一个特点是产生“逃离效应”的危险性比生物进化要大得多。在生物进化的过程中，产生逃离效应的条件是比较苛刻的，因此偏离最佳点的现象在生物进化中并不常见，认为某种性状总是使生物处于适应度的最高点通常不会导致错误结论。但是在文化进化的过程中，引起逃离效应的因素相当普遍，虽然文化价值观归根结底受

到生物价值观的约束，但是文化价值体系偏离最佳点的情况是很常见的。因此对于生物的性状或作为生物性状的生物评价体系来说，我们认为它在几十亿年或几百万年的进化过程中已经作到了充分的优化是大致不会错的，但是文化价值体系却未必有这样高的可靠性，根据某种文化现象已经存在了几千年就认为它当然合理是很危险的。

在文化进化的过程中，存在着多种导致逃离效应的因素，其中一个重要的因素就是从众心理。由于文化价值体系中有很多价值观是对于集团或社会起作用的，当某个个体偏离这种价值观时，尽管这种偏离有可能是更有利的，但偏离本身却使得这些个体对于多数人来说可预测性下降，合作变得困难，与社会结构发生冲突，带来的损失会超过个体价值观向有利方向偏离所带来的利益。这种因素也可以表现为标准化的利益与改良之间的冲突。从众心理是保持文化的稳定性的主要力量，也是文化进化的阻力，这种力量使得文化在一些方面总是落后于现实所要求的价值体系。导致文化进化的逃离效应的另一个重要因素是文化价值观的反生物性。文化价值观独立于生物价值体系的根本性理由在于囚犯困境问题的哈丁解，也就是一个集团和社会在某些问题上总是需要一种外在于个人的权威力量来维护整体的（因此也就是每个个人的）长远利益不被局部而近视的生物价值观所误导。为了实现这种功能，通常以风俗、道德或法律形式表现的文化价值观必须具有足以克制生物价值观的力量，而且几乎总是代表一种与生物价值观相反的评价倾向。因为与生物价值观同方向的文化价值观通常没有存在的必要，我们用不着把每天吃饭看成是一种道德的行为。文化价值观的这种反生物性的方向和力量，使得用生物价值体系对其进行约束变得相当困难。为了保持这种力量，文化价值观往往被涂上神圣的色彩，这也就是宗教产生的根源，而宗教教义从来都是天然地处于免受批判的地位。某些形而上学的价值观

虽然没有打着宗教的旗号但只要它同样被涂上神圣的色彩，就具有和宗教相类似的拒绝批判的性质。对于文化价值体系进行批判和修正的困难还在于，任何一种批判或比较都必须有价值体系作为依据，而处于文化深层的价值体系自身就起到这种作用，从而使自身处于不可能被怀疑的地位。只有在不同的文化之间发生冲突时才可能产生对文化价值观的质疑，但是由于不同的文化之间通常不存在共同的基础，使得有效的争论无从开始，除了各执己见之外别无他法。因此文化价值观的进化通常不是通过争论达到一致的，也不是可以靠通常的逻辑推理解决的，而是靠人的生物价值观所造成的直觉决定取舍的。建立可以作为讨论价值观的依据的价值理论体系需要从更深层的生物价值观入手。

在模块结构的信息系统之中，由于上层模块共用下层模块的可能性，以及动作目的不同所产生的制约，使得上层模块必须避免同时动作，从而使上层模块之间具有互抑制的关系，即一个模块的工作状态导致其他模块的休止状态。而该由哪一个模块工作，可以由模块之间的竞争来实现。即使这些上层模块都由一个中央集权的最高级模块进行集中控制和协调，其控制机制也仍然是竞争的。竞争的胜负总是体现出某种评价，因而模块之间的竞争可以说是评价过程的最基本的表现形式。我们用蚕吃桑叶的简化模型作为具体例子来说明这竞争过程。假定蚕有两个运动模块，一个是咬吃，一个是爬行，在没有桑叶可吃时，咬吃模块处于抑制状态，爬行模块工作使蚕爬到有桑叶的地方，一旦碰到桑叶，某种传感器的信号使咬吃模块起动，爬行模块被抑制，使蚕能吃桑叶，一旦头部摆动范围之内的桑叶被吃完，咬吃模块兴奋度下降，对爬行模块不再抑制，蚕再次爬行以吃到别处的桑叶。从这个过程来看，要使蚕完成必要的生理机能，咬吃模块的评价优先度必需高于爬行模块，如果评价发生反转，爬行的优先度高于咬吃，那么蚕就会永远走下去没机会吃桑叶了，如果真有这样的蚕当然会

饿死而被自然淘汰，因此在进化过程中必然产生合理的也就是适应的评价体系。这个模型还可以进一步复杂化，例如我们考虑蚕需要脱皮，在脱皮的过程中既不吃也不动，那么可以假设还存在一个脱皮模块，其优先度高于咬吃和爬行，在需要脱皮时脱皮模块可以抑制咬吃和爬行模块，使蚕处于不吃不动的状态。在蚕的神经系统中未必真有一个与脱皮有关的神经节，很可能在脱皮时蚕的体内产生某种激素或某种酶，这种化学物质使得蚕的皮肤发生变化，同时又能抑制神经活动。但是作为一个信息模型来说，信息的作用到底是由物理的还是化学的变化来实现并不是一个值得注意的因素。促成脱皮的化学物质到底是一种什么样的激素或什么样的酶，这个问题对于研究杀虫剂的人来说可能事关重大，但对于建立信息模型来说不起关键作用。

蚕的简化模型不包括学习过程，所有的评价都是由先天的本能决定的。对于能够学习的动物来说，有些评价是可以通过学习而改变的。例如前文所述的蜜蜂学习颜色的过程，从评价的角度来看，可以认为蜜蜂通过学习改变了各种颜色对于近距离导航的价值。学习算法可以采用 *ES* 学习的算法。对于价值的学习来说，必需由更深一层的价值作为学习的引导价值，这在 *ES* 学习算法中体现为成败判断的基准，在蜜蜂对颜色的学习中表现为吸蜜成功的高价值评价。因此复杂的评价系统也可以由多次的学习逐层建构起来，其中最深的层次则是由本能决定的基础价值体系，由进化的机制保证其适应性。通过类似的学习过程，认知主体可以对于各种感觉信息进行评价赋值，由这些评价的对比可以象本能价值一样引起价值中断，以控制和协调行为和思考的流程。因此认知主体所接受的外部刺激对认知主体的扰动作用不仅仅取决于刺激强度，更取决于感觉信息所具有的先天的或在过去的学习中获得的评价值。

利于价值中断的概念，我们可以对一些难解的心理现象作出

解释。例如当我们睡着的时候，可能不被雷声吵醒，但有人小声叫名字时却会被叫醒，这是由于在长期生活的学习过程之中，自己的姓名总是与对自己有意义的事相连，这一信息被赋予了很高的价值，足以对睡眠状态进行中断，而雷声却往往不具备这种价值。但是如果你住的房子漏雨，雷声作为与雨有关的信息也会在学习中获得较高的价值。应用相同的原理也可以解释被称为“鸡尾酒会效应”的心理学现象。所谓“鸡尾酒会效应”是指这样一种现象：在鸡尾酒会上，你和朋友正在谈话，其他人也各自在谈各自的话题，你只听到朋友说的话，其他人的谈话则构成“背景噪音”，但是当你近处的某个人在谈话中提到你的名字时，你会清楚地听到，而且引起注意。为了分析这个现象首先需要假定我们的听觉信息处理处在较低的层次上（例如单词或短语这个层次）是可以并行处理的，即可以同时处理两组以上的信息，但是高层次的语句的语义理解由于需要智能模型的参与，只能处理一个语流。在主通道进行语言的全面信息处理的同时，副通道只需要作出词语的包括重要性评价和单词语义的简单处理。通常主通道抑制副通道的高层次语义处理，但是当副通道的评价值超过主通道时，会发生价值中断，使注意点及高层次语义处理切换到原来的副通道上，造成主副通道的互换。自己的名字具有很高的评价价值，足以触发这样的切换。当然，价值评价是相对的，如果你和朋友所谈的问题极为重要，中断也可能不发生，反之，如果你正在被通缉，一个可疑的脚步声也可能引起中断。价值中断在多数情况下可以在短时间恢复，中断引起主通道切换之后，如果主通道中更高层的处理所得到的价值低于切换前主通道的信息价值，主副通道可以再次互换使中断得到恢复。鸡尾酒会上由于名字引起的中断，如果听下去之后并无吸引人的信息，和朋友的谈话又会继续下去。

通过学习对感觉信息进行评价赋值，使得感觉信息带有评价价值，不仅在与外部世界的信息交换过程中起作用，而且对于思考

过程也起着引导作用。思考是动作的内化，思考的每一步的结果造成内部的感觉，由于这些感觉信息也同样在过去的学习过程中获得了评价赋值，所以可以作为思考推进的依据，当思考得到的感觉信息评价价值较高时，思考可以集中于原路，当评价价值下降时，思考会走向散漫，发生下意识的分路进行，准备在某一路上出现评价价值较高的感觉信息时，再次集中思路。这种过程如果看成一个搜索过程，感觉信息的评价价值就相当于启发信息。以上的分析为我们开辟了通过学习获得启发信息的途径。

信息的评价价值这一概念，为建立新的信息理论开辟了方向。信息论的创始人香农（C. E. Shannon）从统计学的角度对信息量作出了严密的数学定义，阐明了信息与熵的关系，使得信息论成为一个坚实的理论，对于解决信息传输和编码等许多问题提供了可靠的工具，在信息处理中也起到指导作用。信息量这一概念的优点是其源于客观性的确实性，因此对于象信息传输这种客观的过程的研究和设计具有非常可靠的指导意义，但是单纯的传输本身并不使信息具有意义，信息只有在和认知主体发生相互作用时才体现其意义，因此更完整的信息理论应该包含信息与认知主体之间的作用。例如你的妻子（或你自己）生了一个男孩，与你不小心掉在地上的硬币正面朝上，这两个事件所代表的信息计算其信息量是完全相同的。如果排除人的主观感觉，一本无聊的厚书比一本精彩的薄书信息量要大，甚至同一内容的书有错别字的比没有错别字的信息量要大。信息量概念的这种明显的不合理性是早已被人们所熟知的，有不少人主张建立新的信息理论来解决这种不合理性，但是都没有成功。其原因在于历来的科学传统是重视客观性和确定性的，由于主观评价的不确定性，不可能被包容在一个可靠的理论之中。感觉信息的评价价值概念，在确定的学习算法之下，是具有其确定性的，至少具有统计意义上的确定性。按照这一思路，我们可以建立关于主观评价的具有确定性的或者说

具有客观性的理论。由此就可以建立起包含信息的发出者和接收者的主观评价因素在内的，同时又是科学而可靠的广义信息论。

以上关于感觉信息评价值的讨论只考虑了评价值在竞争中的相对比较，而没有考虑感觉信息评价值在不同心理状况下取值的变化。为了讨论这种变化，我们可以不把评价值看成是某个感觉信息所固有的，而是感觉模式所具有的评价子向量与代表认知主体现实感情状态的感情向量之间相互作用而产生的评价值。这种即时的评价作用可以由具体的算法来实现，例如我们可以把感觉信息的即时评价值定义为该感觉信息的评价子向量与当时整体感情向量的内积。反之，对于整体感情向量对于某种感觉信息的评价子向量的影响，则可以采用和 *ES* 学习算法相似的方法，只是把调整概率改为调整评价子向量各元素的值。调整可以用以下的算法实现：学习后的评价子向量是学习前评价子向量与以成败作为正负号的即时感情向量的加权平均值。通过这样的相互作用，感情评价可以起到需要哪种感觉时哪种感觉的评价值就较高；而与成败无关或导致失败的感觉信息获得较低评价的引导作用。这种感情评价的引导作用比起模块之间直接的互抑制或互兴奋具有更高的可塑性和对环境变化的即时响应及变通能力。

对于行为方式固定而简单的动物来说，行为模块之间固定的优先度排序已经足以应付评价的需要。例如在蚕吃桑叶的例子中，蚕在没有受到中断情况下只管一路吃下去，这种行为方式所表现的永恒的食欲在行为机制内部并不需要食欲的概念。食欲只有在可能发生变化和可能与其他欲望进行比较和竞争时才能成为有意义的概念。对于行为方式复杂而且可以通过学习加以改变的动物来说，欲望作为某一类行为的代表进行各类行为之间的比较和竞争就具有简化竞争方式的重要意义。从学习的角度来看，象蜜蜂学习颜色或乌鸦区别对象这种依付于固定行为方式的目标识别学习，以及象马戏团驯兽这样的出于固定目的的行为学习，由于键

刺激与行为之间形成固定关连，而学习所依据的一方已由本能固定，欲望的作用也不太大。对于象人这样在诸多方面具有复杂而多层次的学习能力，往往在 R 和 S 两方面都要进行学习，学习所得目标识别和行为之间的关系又可随环境因素而变化的认知主体来说，欲望这种相对稳定的评价基础就有其极为重要的意义了。欲望反映了人类先天本能的基础评价体系，虽然后天的学习过程和所处的文化环境可以使其发展、对其进行修饰或抑制，但难以完全消灭。在 S 型学习和 R 型学习交互完善的复杂学习过程中，欲望起着始终一贯的评价支持作用。例如婴儿在通过 R 学习过程学会抓东西放在嘴里，由此就可以通过味觉经验学会从视觉和触觉分辨可吃和不可吃的物体，从而产生从欲望到目标的联想，再后又逐步学习获得目标物以满足食欲的丰富多彩的行为方式，形成从欲望到目标的联想，学会由目标起动行为，进而由各种行为经验反馈出目标之间的评价，以及评价值与环境的关系等等，所有这些复杂的学习过程都需要由食欲这个统一而稳定的评价基础作为其先决条件。在通过多层次的反复学习建立起从欲望到目标和从目标到行为的联想能力，以及各种目标及行为方式在环境条件下的评价系统之后，就可以在欲望产生时利用这个评价系统找出满足欲望的最佳行为方式。在学习目标和行为之前，欲望成为学习的动力和导向，动力体现在产生不满足感，中断其他欲望产生的行为，在情绪上表现出焦躁不安，促进认知主体对环境进行搜索，增大发现目标的概率；导向则体现在设定整体感情状态向量处于与欲望相符的取值，使得能让欲望得到满足或与欲望有关的感觉信息能够获得较高的评价，成为发现目标的引导信息。因此在我们饥饿时会产生空腹感，能够对食物的气味产生比平时更为灵敏的感觉。在经过对目标的学习之后，食欲往往会引起对各种具体的食品的形象及味道的想象。

欲望是具有复杂学习能力的动物实现学习与行为的基本动

因，起着设定整体感情状态向量的作用，各种欲望之间的竞争结果，最终决定认知主体的学习与行为。但是在我们的带有感情评价色彩的感觉信息之中，有些感觉并不一定只与某一种欲望有关，有些则很难归属于某种欲望，因为欲望这个词在语义上带有积极的意义，体现某种主动追求的倾向，所关联的多是一些正向的评价，而在感觉信息中还大量存在着负向的约束性的评价，体现出对某种因素的逃避倾向。如果硬要造一个在这种情况下与欲望相对应的词汇，也可以称之为某种“负欲望”。但是这个词用起来很困难，因为很难从日常生活中找出属于“负欲望”的例子，我们熟知食欲、性欲、求知欲这种表现正向欲望的概念，却没有象“避痛欲”这种表现负向欲望的词汇。尊重我们的语言习惯，对于负向的一些评价因素可以不从欲望出发而分别讨论具体的带有评价色彩的感觉信息。对于无法归属于某种欲望的感觉，可以从评价的更深层的基本原则来加以讨论，而欲望本身的合理性也是由于符合这些基本原则，即我们在第二章所讨论过的存在原则、经济原则和预测原则，不言而喻，食欲和性欲是符合存在(生存)原则的。

存在原则不仅体现在食欲和性欲这种趋利型的正评价之中，更多地体现于保护肌体不被破坏的避害型负评价。避害型负评价的各种因素之间竞争的评价优先度是根据“轻重缓急”原则排列的，轻重是指对机体的危害程度，缓急则是造成危害所需的时间长短。这些固有的负评价值在进化过程之中趋于合理，为动物或认知主体指示出为了生存在当前需要解决的问题。饥渴、缺氧、外伤都会威胁动物的生存，但缓急程度不同，饥渴的时限是数天，窒息是数十秒，外伤则要求瞬时的反应。因此，一般来说，造成外伤的高温、外力的负评价值大于窒息，而窒息大于饥渴。实际的中断竞争则取决于不利刺激的强度，而具有经时效果的不利刺激的强度则随其持续时间的增强而增大，在接近时限则急剧增加。

在存在原则所支持的诸多保护性、避害性的负评价感觉之中，

具有代表性的是疼痛感觉。痛觉可以由压迫、撞击、刺切、感染、化学腐蚀等多种破坏性因素所引起，是避害性感觉中最典型的例子。由于痛觉的存在，使人生增加了一些痛苦，但是痛觉却是人所不可缺少的一种感觉。曾经见到过先天缺失痛觉的病例的报告，在其他人的看来，没有痛觉似乎是人生一大幸事，可以免去很多痛苦，但是对患者来说，却造成了他人难以想象的麻烦。由于没有痛觉，患者不能象正常人那样及时获得肌体被破坏或可能将被破坏及损害程度的信息，这样就会使动作学习免受损害的功能难以发生作用，因而患者经常被碰得伤痕累累，即使在事后可以发现自己的伤痕，但却无法得知是在哪一个动作过程中碰伤的，因此对动作的修正变得相当困难。痛觉是一种主观感觉，压力、内部接触、高温传感器都不能直接作为“痛觉传感器”使用，作为痛觉必须加上对损害程度评价的翻译过程。这种翻译过程并不是对所有动物都是必要的。对于简单的动物来说，避害动作是由基因决定的神经互联来完成的，某种对动物不利的刺激直接引起相应的逃避行为，如蟑螂对振动的敏感反应。在这样的行为模式之中，痛觉这个中间环节毫无存在的必要。对于固定的单一因素的学习过程，例如蜜蜂对颜色的学习来说，颜色评价只与近距离导航有关，并不与其他行为发生竞争，因而关于快感和不快感之类的感觉也是没有存在必要的。只有在具有复杂的学习能力，需要广泛而多变的评价竞争的认知主体之中，象痛觉这种具有通用评价可能性的感觉才是有必要的。从这一点出发，我们可以推断节肢动物是没有疼痛感觉的。据报载澳大利亚颁布了一条法律，由于龙虾具有神经系统，为了减少龙虾的痛苦，在享调之前必须弄死，禁止活杀或活煮。且不谈造成动物的痛苦是否违反道德的问题（对于这一点孔子已有一个很精采的解决方案：君子远庖厨），即使承认减少痛苦的必要性，龙虾这种节肢动物也没有资格成为考虑的对象。龙虾在热锅里乱爬仅仅是直接的动作反应，并不存在疼痛

这种中间环节。就象我们在第一章中谈到的挨打机器一样，它没有也不需要疼痛的感觉。

快与不快的评价性感觉在认知主体中有两个主要作用，一个是作为学习中对行为进行修正的引导信息，一个是作为将来的行为决策即块竞争的评价信息。洛伦兹(Von KONRAD LORENZ)把这种学习与决策的过程比喻为法庭的判决，但是这种比喻只适合于比较简单的动物，对于更复杂的动物，用市场作为比喻可能更为恰当，法庭的比喻适合于中央集权的系统，而市场则是分散自律系统，法庭从理论上依靠不变的法律进行判决，而市场上的价格和汇率总是不断变动的，法庭不包含立法机构，而市场可以自发产生市场规则。对于简单动物来说，模块之间的优先度顺序是由进化过程赋值并固定于基因中的，其行为方式也是由基因决定程序完成，不能由学习加以改变，在这种情况下，行为的启动近乎程序安排，而行为的完成也只需要给出完成信息以结束行为即可，完成信息之中不需要包括评价成分。而对于具有学习能力的认知主体来说，不仅需要保护性的不快感觉（例如痛觉）来即时约束不正确的动作，而且需要行为成功时的快感以巩固正确的行为，这种对行为结果的评价也需要记忆起来作为今后与其他行为竞争决策的评价赋值。对于一生只交配一次的昆虫来说，其交配行为方式只能由基因决定，不可能借助经验进行学习，而且也不需要未来的竞争参考，因此在这样的动物的交配过程中，观察不到情绪高昂的表现，看起来像完全程序化的例行公事。对于具有高度而复杂的学习能力的动物来说，情况完全不同。本能规定的行为方式减少，而作为学习引导的评价功能则高度发达，基因不再完全规定具体的行为方式，而改为规定高层的欲望。把两只性成熟的猩猩放在一起，它们从未见过其他猩猩的交配过程，结果猩猩虽然表现出极度的兴奋，却不知如何解决这个问题，通过试错的方法它们或许可以发现性行为的动作，但是在没有示范的

情况下，恐怕需要较长的过程。

从上述观点出发，我们就可以回答那个古老的问题“子非吾，安知吾不知鱼之乐？”鱼的游泳动作是一出世就会的，而且不会通过学习改变游泳的姿势，因此对于鱼来说，游泳时判断那种姿势更好是没有意义的，当然也不会有游泳时的痛苦与愉快之分。观鱼者认为鱼游得快乐，仅仅是因为鱼游动时动作频率较高，而人又是在快乐时容易有较高频率的动作出现，从而作出鱼游得很快乐的猜想，而作为鱼的自身感觉来说，并不需要游泳时的苦乐之感。当然这是对鱼而言，如果所讨论的不是鱼而是海豚，情况就大不一样了，因为海豚是很善于学习新动作的动物。这个古老问题的深刻意义在于，在看到行为主义的缺陷之后，我们用什么方法去判断除了我们自己之外的认知主体（不论是动物还是机器）是否能思维，是否有主观感觉，是否有感情，是否有意识。所有这一切，都取决于我们能否对这些通常表现主观世界的词汇给出客观的模型和阐明其机制，从这一点出发，我们就能通过结构和动作原理从客观的角度分析一个认知主体的主观世界。

体现经济原则的约束性负价值评价感觉的典型例子是疲劳感。从生物化学的角度看，疲劳产生的机制是肌肉运动时乳酸和二氧化碳之类的代谢产物来不及随血液排出而积累起来，使肌肉酸度增高，酸度由相应的感觉神经检测出来传递到大脑，从而产生疲劳的感觉。疲劳感可以起到一定程度的保护作用，走路走累了坐下休息一下，可以使积累的代谢产物有时间排出，恢复肌肉的正常功能。然而，疲劳感觉的意义远不止这些，疲劳感在不同层次的学习和决策之中起着极为重要的作用，是认知主体对于行为所付出的能量代价的度量。按照经济原则，动物行为所付出的能量越小越好，因此疲劳感带有负的评价价值。在低层次的动作学习之中，疲劳感是对动作进行优化的关键信息。在某种技能的熟练过程中，或是动作的反复练习之中包含着一种渐近性的学习过

程，每一次的动作都有小的随机变化，而用疲劳这种反馈信息决定动作修正的方向，以达到在完成相同目的时所付出的能量代价最小这一原则。熟练的动作往往比较圆滑，不象简单的机器人那样作折线运动，如果给机器人加上加速度最小化的学习程序，则机器人的动作也可以变得圆滑起来，而对动物的肌肉来说加速度与力即肌肉紧张程度成正比，而肌肉紧张要消费能量。在动物或人学会某种基本技能之后，熟练的最重要任务就是消除不必要的肌肉紧张，因为在掌握基本技能时需要积极的多方向搜索，多余的紧张是不可避免的。这就象一个新产品进行原理性试验时可以不计成本，但要投入大量生产时成本就成为重要因素了。在降低运动成本的学习过程中，疲劳就成为重要的信息了。刚学会骑自行车的人，长时间骑车会感到手臂很累，这是由于学习平衡动作而产生的过度紧张还没有消除，经过一段时间的练习之后，在疲劳信息的指引下，过度紧张就会消除掉。初学游泳的人，只能游几十米的时候，还不能算会游泳，此时基本动作虽然差不多，但用力还很不得法，游几十米体力就支持不住了，到能游几百米时，才可以认为已经基本上掌握了正确的用力要领。疲劳状态本身也促使紧张的消除和多余用力的减退，迫使搜索向减少用力的方向进行。当然经济原则不是只省力，而是效果与成本之比，因此在对支出作出负评价的同时，应对效果作出正评价。在游泳的例子中，速度感是带有正评价的感觉。只有在正负评价的共同作用下，动作才能实现经济意义上的优化。对人来说，速度的正评价是比较普遍的，开汽车兜风的乐趣也是这种正评价所产生的快感。从决策的角度来看，疼痛感的记忆可以使未来的决策避免导致伤害的行为，而疲劳感的记忆则使决策能正确估计能量代价，避免不经济的行为。我们所说的“懒得”干某件事情，就是通过代价与效果的评价所得到的否定性结果的语言表现。

体现预测原则的一些评价性感觉不象疼痛和疲劳那样明显，

但在复杂的认知过程中的重要性却不亚于前两个原则。好奇心和求知欲可以说是预测原则的集中体现。关于脊椎动物眼睛的分布与预测能力的关系，洛伦兹的研究很有启发意义。脊椎动物的两只眼睛的分布有两种类型，一种是两侧分布的，其视野较宽阔，但两眼的共同视野较小，另一种是两眼都朝前方的，总视野较小，但两眼视野的重合部分很大。大多数的鱼类的眼睛都是两侧分布的，这些鱼类对环境的纵深距离感觉主要依靠游动时视网膜中景物随移动而产生的变化来获得，对于这些鱼来说，运动过程中的随时的反馈足以使其避开障碍，游向目标。但这些鱼在静止时难以得到足够的空间信息，因此会有这样的情况发生，当鱼在前方有障碍而处于静止状态时，先向前方游动之后才作出躲避障碍的动作。有一些鱼类处于向陆生动物进化的中途，浮漂已经退化，不能在水中游动，只能在水底或水边的岩石或沙滩上生存，靠跳跃来移动身体。对于这种鱼来说，跳跃之前总是处于静止状态，为了准确地跳跃，必须能在跳跃之前正确测定距离以决定跳跃的力量，因此这种鱼进化成两眼朝前方的分布，可以在静止时不依靠运动视差而只靠两眼视差来测定距离。动物就是通过这样的进化过程获得了初级的预测能力。在鸟类中也有类似的例子，大多数的鸟的眼睛都是两侧分布的，只有猫头鹰是个例外。普通的鹰是在飞翔过程中发现猎物的，因而运动视差足以满足测距要求。但猫头鹰要抓老鼠，成为夜行性动物，而在夜间听觉比视觉更为有效，因而猫头鹰长了两个羽毛构成的耳朵，由于飞翔中气流所引起的背景噪音太大，为了听取猎物的声音，猫头鹰只好站在树上发现猎物。为了能在静止状态下正确判断距离，猫头鹰的眼睛是双眼朝前分布的。在哺乳动物之中，猫是双眼朝前分布的典型，因为猫的捕食方式是在静止埋伏的状态下突然扑向猎物的，双眼视差的测距作用对生存有重大意义。猴子在树上从一个树枝跳跃到另一个树枝时要不失时机地握住目标的树枝，而且在飞跃过程中无法

控制自身轨迹，必须在跳跃之前的静止状态下正确测量距离，因此灵长目是两眼朝前分布的。人类的高度预测能力就是起源于此。

预测原则在认知主体的复杂逐层学习的过程中也是不可缺少的。体现存在原则的保护性负评价只起约束作用，不起正向引导作用，而只靠经济原则的效果作为正向引导的话，效果的表现又成为一个问题。特别是在低层次的学习过程中，行为通常不引起直接的利益，此时预测原则成为最重要的学习引导的正向评价信息，它使得通过学习建立的智能模型具有预测能力。这就是为什么我们通常用胡萝卜加大棒的政策训练动物，但没有必要在小学课堂上发糖果。预测原则有别于存在原则和经济原则之处在于其非短期功利的导向，这就为人类的认知超越动物认知开拓了道路，当然从长远和整体来看，预测原则仍然是合乎功利主义的，这样才能解释预测原则在进化中为什么能够产生而不被淘汰。求知欲在进化过程中被加强，显然是缘于其对适应度的提高，幼年的猴子乐于学习，但随着年龄的增长其求知欲逐步下降，这是由于年青时学习到的技能可以用很多年，学习所花的代价可以从一生的利益中得到充分补偿，但是对于老年的猴子来说，由于其后的时间不多，学习新技能可能得不偿失。随着智力与文化协同进化的发展，学习的利益越来越大，人保持求知欲的年龄也逐步增长，趋向于幼年成熟，因此人的头盖骨形状与幼年猴子的头盖骨极为相似。虽然出于同样的价值核算人也是年纪越大越趋于保守，但与猴子相比，人可以算是终生学习的动物，其中不乏象科学家这样的活到老学到老的例子，因此可以说，人是永远长不大的猴子，而科学家则是永远长不大的人。由此可知人的生涯是有其特殊性的，根据其他哺乳动物性成熟期与寿命的普遍关系来推断人的寿命应为150岁可能是不对的，人的幼年成熟的特性可能推迟了性成熟。

求知欲主要表现在好奇心和保守性两个方面。好奇心是对未知的或者说是不可预测事物的关心和兴趣，保守性则是对事物可

预测性的渴求和对新事物用原有智能模型进行解释的倾向。对事物进行理解以增加知识的过程大致可以分为两个阶段，在遇到输入信息时先是驱动智能模型进行预测，如果预测成功而且所得结果并不重要，则不引起价值中断，如果预测成功而结果重要，则引起中断并进行进一步的思考决策，如果预测失败则也引起中断，换用非典型的其他模型试行预测，视其成败改变典型模型分配，或通过进一步思考修正模型本身。好奇心也就是对预测失败给出比较大的价值的绝对值，使认知主体通过学习改变智能模型，而改变后得到的智能模型应能在更大范围得到正确的预测，这是保守性的要求。对于婴幼儿来说，整个环境都是未知的、新奇的，作为认知主体内部的智能模型还很少，此时认知主体通过对具体事物的记忆得到智能模型，然后利用这些模型进行预测，对预测有效的模型得到保留。无效的被淘汰。在建立了一些有效的智能模型之后，逐步在进一步的预测理解的过程中推广模型的使用范围，并通过建立模型之间的关系组合成更高层次的智能模型。在这些过程中，预测失败的中断是起动思考或行为尝试建模的重要信息，其表现就是好奇心，所谓“奇”就是用认知主体中原有的信息模型难以解释即难以预测的信息。

但是人并不是对所有自己不懂的事情都感兴趣，好奇心仍然受到短桥原理的约束。人所感兴趣的，是那些对智能模型稍加修改或简单组合就可以解释的问题。但是在多次的修改和逐层的组合之中，智能模型可以逐步完善和向高层次发展。具有一定语言理解能力的儿童，喜欢旁听成年人的谈话，虽然他们还不能理解谈话的全部内容和深层的意义，但儿童可以在这些话中选择在他当时的水平上经过思考可以理解的部分加以理解，同时逐步改进智能模型，扩大理解的范围，最终达到完全理解的程度。这种选择所依据的原则就是短桥原理，非常容易理解的部分不引起中断思考，太难以理解的部分则引起理解的失败或错误，这些错误有

待将来理解能力进一步提高之后加以修正，起作用的是那些在短桥距离之内的内容，即通过思考可以完善和提高智能模型的信息。这就是容易引起人们兴趣的信息。具备初步外语水平，然后生活在外语环境之中的人，可以经历类似的学习过程，在一开始只能依靠几个单词去推断语句的意义。根据语句的意义再去推断其他单词的含义，逐步提高语言能力，最后达到能够理解话外之音的程度。从基于预测原则的评价的角度看，可预测性太高的信息评价较低，我们看一部不久前看过的电影或是废话连篇的电视剧会感到无聊。反之，新奇性太高而超过短桥距离的内容也会引起不快的感觉，例如过于离奇的故事或超越自身理解水平的讲课，以及别人用完全听不懂的外语讲话，都不能引起我们的兴趣，因为这些信息无助于智能模型的增长。

感情评价的外部表现就是表情。人类表情的丰富程度远远超过其他动物，但表情的生成与识别仍是以本能为依据的。婴儿从一生下来就会用哭与笑表达其快与不快的评价感觉，不同民族的不同文化并不会使表情产生相反的理解，用点头表示否定而用摇头表示肯定的民族是存在的，因为点头摇头是由文化所规定的信号，但是用笑表示悲伤，用哭表示高兴的民族是不存在的，因为表情受本能的控制，不因文化而改变其表现形式。因此表情要伪装得很象也是很困难的，假笑与真笑很容易分辨。这也就是斯坦尼斯拉夫斯基体系的理论根据。人的表情如此丰富，也是产生于智力与文化的协同进化过程。随着文化对于生存的重要性逐步增大，文化的高效率传递方式对于提高适应度具有越来越明显的效果，而文化的核心部分又是该文化的价值观体系，因此表情的生成与识别的能力作为传递非本能文化价值观的有效方法在进化过程中成为一种适应的本能而得到保存和发展。

人类表情的典型例子是笑。笑又分为微笑与大笑两种，表现不同的感情。微笑产生于愉快的心情，进而对他人表示赞许和友

好。大笑则是对新奇的或超乎常态的事物或说法的反应。司空见惯的事不会引起大笑，引起大笑的事情多少有些怪异，因此大笑可以说是在短桥距离之内的新奇事物引起的中断状态。大笑总是中断平常的思路，对引起大笑的原因造成反复回忆，增加记忆的深度，达到学习的目的。

在前面的讨论之中我们没有对感情与评价作明确的区分，实际上在日常语言或心理学中所说的感情通常是指可以被意识到的感情状态，而无意识之中的评价过程通常不被认为是一种感情状态。严格地说，评价是感情的基础，感情是被意识到的评价结果的表现。除了有意识的决策过程之外，评价过程本身通常是不被意识的，但评价的结果会引起身体和思考状态的变化，例如心率的变化，表情的变化，激素分泌的变化等。这些变化的作用本来是使动物为评价决策之后的行为作出充分的心理和生理的准备，或是生成表情信息或语言信息向外部进行表现以影响其他个体。评价结果的外部表现是一切社会性动物的必要功能，例如蜜蜂对蜜源或分封时筑巢地点的评价结果可以用舞蹈的方式加以表现，狗也可以用摇尾巴表现其好感，争斗向仪式化方向的进化也可以形成评价结果的外部表现。评价结果所引起的身体状态变化和外部表现通过感觉被意识到时，就形成我们通常所说的感情。因此感情是与意识密切相关的概念，要理解感情的作用需要以对意识的理解作为基础，这里只能先指出一点，即感情作为评价进入意识的途径，使得对评价系统的“事后”的修正成为可能。

4.7 意识与自我

意识是认知科学之中极为重要但又最为困难的问题。关于意识问题的重要性，现在已经引起认知科学界的充分重视，诺尔曼把意识列为认知科学的12大课题之一，但是关于意识的研究，目

前还只停留在现象的水平上，而缺乏一个有说服力的关于意识的理论。19 世纪的心理学研究主要集中于意识到的心理学现象，当弗洛伊德提出无意识领域的存在时，曾受到不少心理学家的反对，但是到了 20 世纪，由于行为主义的实验心理学派的发展，与意识有关的心理学研究反而被看作形而上学而受到排斥，直到近二三十年以来，行为主义的局限性逐渐暴露出来，与意识有关的研究又重新受到重视。

关于意识的研究在认知科学中的重要性虽然已经没有什么人表示怀疑，但是对于意识本身还存在着很多不同的看法，而这些看法的分歧主要表现在接近哲学的层次上。关于这些分歧，P. N. Johnson-Laird 在《*Mental Models* (1983)》一书中作了一个简单的总结：“没有人真正知道意识是什么，意识起什么作用，完成哪些功能。但是可以提出一些二者择一的问题，就是：意识是完全不可理解的东西，还是可能用科学的方法加以说明的东西；一切可实现的手续都是可计算的这个丘奇—图灵假说是正确的，还是，意识是一种未知的，不可计算而又可实现的手续；功能主义的主张是正确的呢，还是精神不仅立足于大脑的活动，而且依存于大脑的特定的物理性质。根据对以上问题的不同回答，意识应该属于以下四个范畴之一。

1. 超自然现象。也就是在意识机中存在着用科学无法说明的‘灵魂’。（灵魂说）

2. 意识可以用未知的不可计算而可实现的手续加以说明。（不可计算说）

3. 意识象天气一样，在原则上可以给出一个能用计算机程序加以模拟的理论。但是和实际上在计算机中并不可能实现高压一样，意识也不可能在计算机中实现。只有具有大脑的生物才有可能具有意识。（模拟非真说）

4. 意识是可计算的。给计算机编一个具有意识的程序在原理

上是可能的。(可计算说)”

以上四种说法的第1种说法即灵魂说显然不能引起科学家的兴趣和赞同，因为接受这种理论意味着把意识排斥于科学研究的领域之外，或者说把科学家全部排斥在意识研究的领域之外，成为研究意识的科学家的自我否定，不符合存在原则。第2个说法即不可计算说也有相似的缺点，要么毫无根据的接受这一观点从而拒绝对意识的可计算模型的研究，要么也只能通过研究意识的可计算模型的过程来证实其不可计算，因此不可计算说作为研究的出发点是毫无意义的。第3种说法即模拟非真说看起来有点道理，也就是说在计算机中用程序模拟的暴风雨的确不会把计算机淋湿，但这是在我们把暴风雨看成物理过程时才如此，计算机模拟的是物理过程中的信息的侧面。站在信息的角度上看，情况就不一样了，例如我们可以用一个微处理芯片去模拟一个气动的PID调节器，但此时我们不能说这个由程序动作的微处理芯片不是真正的PID调节器，因为从信息处理的角度上看二者作为调节器是完全等价的，微处理器里有没有气动仪表的波纹管中的气压是无关紧要的。这是维纳的《控制论》所建立的一个基本观点。只要把意识看成是一种信息过程，那么这个信息过程不论是发生在人脑中还是发生在硅元件中，都没有真假之分。把意识看成是物理过程，或是生物所特有的“活力”产生的过程，或是超科学的“灵魂”作用，都比看成是信息过程更加令人难以接受。因此第4种说法即可计算说是唯一可以作为意识研究的出发点的基本立场。

关于意识的研究成果大多局限于现象的水平上，但这些研究所得到的结论对于建立关于意识的理论模型是不可缺少的，任何关于意识的有效的理论模型都应该对于意识的现象具备解释和说明的能力，因此意识现象就构成理论的约束和检验。关于意识的已知的重要性质有以下几点：意识的串行性，意识的内省能力，意

识的主动性，意识的控制（选择）能力等等。任何现象都没有直接通向理论的必然途径，理论只能靠思辨的方法获得，然后经受现象的检验。作为寻求意识理论的线索，我们应该回答这样的问题：意识在认知过程中起什么作用。回答了这样的问题之后，我们就可以根据意识所起的作用去分析其功能，进一步建立为实现这些功能所需要的结构模型。

已经有人提出了一些关于意识的理论，但是这些理论都含有致命的缺点，下面我们分析几个这种理论的例子，以了解关于意识的理论的研究状况。波普尔(Karl R. Popper)和艾克尔斯(John C. Eccles)在1977年合著的《自我及其大脑》(*The Self and Its Brain*)一书中提出了“脑—意识相互作用”理论(简称B—MI理论);引起很大反响。该理论认为意识或自觉的精神是独立于大脑中信息活动之外的另一个世界，认为“大脑皮层中的活动对于主体的自觉经验来说是必要的”但并不能说是“充分的”，“自觉的精神有一种并不完全依赖于大脑事件的特性和活动，这些大脑事件是处在由自觉精神的决定和控制影响下的。”按照丁润生著《现代思维科学》所述，我国哲学界大都不同意艾克尔斯的“B—MI论”，认为他把自我意识看成完全脱离大脑的实体，是在走向调合科学和宗教的关系。这种看法出于中国文化的非宗教深层精神结构，对于西方人来说未必有说服力，在西方人看来“调合科学和宗教的关系”可能算不得什么罪名。站在科学的立场上看，这种二元论的理论存在着致命的弱点。第一是其二元论无法解决从笛卡尔以来的所有身心二元论的共同问题，即物质的大脑与非物质的精神之间靠什么来相互作用。在这一点上作为科学家的艾克尔斯比作为哲学家的波普尔更感到尴尬，他极力想象大脑中的某些柱状模块是“开放的”或半开放的，可以通过微小的信号与精神世界发生“相互作用”，但是这丝毫也解决不了问题，再弱的信号也必须具有物理学基础，无法与非物质的世界相互作用。第二个

弱点是关于意识与精神的非物质解释具有和灵魂说相同的非科学性，作为哲学家的波普尔在这一点上并不感到困惑，他明确地称自己的理论是“机器中的幽灵”，并且认为意识是科学永远无法解决的问题。第三个弱点是这种 *B—MI* 理论把精神和意识看作一个浑然无分析的整体，因而这种理论对于意识的理解起不到任何积极作用。因此这个理论所主张的实际上是放弃对意识进行科学研究，这样的理论当然不会受到研究者的欢迎，也不可能导致关于意识的知识增加。

另一种 19 世纪的关于意识的理论是“伴随现象论”。这种理论认为意识是大脑中物质活动的伴随现象，意识对于大脑活动并不起任何作用。这种观点与行为主义具有类似的缺点，即否定了在对认知的研究中意识的重要意义，白白丢掉了有用的线索。如果不放弃关于意识的研究，而又以伴随现象论作为出发点，那么就需要对大脑活动中伴随意识现象的部分与不伴随意识现象的部分加以区分，并说明这些大脑的信息活动怎样引起了意识这种伴随现象。但是这样一种研究方针与可计算说的研究方针也就没有多大区别了。因为从二十世纪的信息科学的眼光来看，如果某种大脑活动造成了意识这种伴随现象的话，我们只把它看成是在不同层次上的现象说明，“伴随”二字是多余的。正如我们不说软件的运行是硬件工作的伴随现象，不说 BASIC 语言程序的运行是机械语程序运行的伴随现象一样。

关于意识的可计算模型目前还很少，但也可以举出一个例子，就是明斯基 (M. Minsky) 在《*The Society of Mind*》中提出的 B 脑模型。这个模型把人脑分为 A 脑和 B 脑两部分，A 脑执行普通的信息处理任务，与客观世界直接联系，而 B 脑不与外界联系，只与 A 脑联系，起着对 A 脑进行监控的作用。当 A 脑进入非正常工作状态，例如发生混乱或陷入无限循环，B 脑能识别这种非正常状态，而把 A 脑从这种状态中解脱出来，使 A 脑恢复正常工作状态。

这个模型可以解释意识的某种程度的内省的作用，或者说是自我监督的作用，但是作为解释更多的意识现象的理论，这个模型就显得过于简单而不够用了。但是这个模型用来解释霍夫斯塔特(D. R. Hofstadter)在《*Metamagical Themas*》一书中提到的意识的“反寄生蜂性”是足够的。反寄生蜂性的说法来源于关于寄生蜂的一个实验结果。寄生蜂在繁殖时先在地上挖洞，然后捕捉其他昆虫的幼虫，麻醉之后放入洞中，在幼虫旁产卵，待寄生蜂的卵孵化之后，寄生蜂幼虫吃被麻醉的幼虫成长。这种有目的的行为给人一种有远见思虑的印象，然而生物学家的实验打破了这种印象。在寄生蜂捕到幼虫拖回洞时，先把幼虫放在洞口处一定的位置，然后检查洞内无异常，再把幼虫拖到洞里去。这时候观察者趁寄生蜂进洞检查之机把幼虫移动几厘米，寄生蜂出来后并不直接把幼虫拖进洞，而是放在洞口的位置上，再次检查洞内，观察者用这种方法曾经让这个检查动作重复了40次(估计最后是观察者不耐烦了)。这说明寄生蜂的行为是靠一个死硬的程序驱动的，一旦环境变化，就会陷入无限循环而不能自拔，象一张滑了槽的唱片总在重复一个小节。霍夫斯塔特把这种机械性称之为“寄生蜂性”，从而认为意识的重要性质就是“反寄生蜂性”，也就是表现得不那么机械，可以知道洞已经检查过了，这次可以直接把幼虫拖进洞里去。如果仅仅把意识看成是反寄生蜂性的话，B脑模型已经足以解释意识了，B脑的存在完全可以避免任何循环过程。但是就我们一般对意识的印象来说，意识的作用远不只是防止循环。象哺乳类或鸟类这样一些具有初步学习和记忆能力的动物，都具有一定程度的反寄生蜂性，但是由此就说这些动物具有意识，则还显得有些证据不足。某些操作系统也具有中断用户程序循环的功能，但不能说它是有意识的。况且，防止循环也不一定需要B脑，只要在有关的神经细胞中引进某种疲劳机制，循环是可以避免的。B脑理论的另一个缺陷是，如果B脑的功能非常简

单，则无法完成意识的复杂功能，例如如果不分青红皂白对所有循环过程进行中断，人就不能走路，鸟也不能飞了，走路当然对应着神经回路的循环。而如果 B 脑本身具有复杂的功能，那么它本身也就容易出现异常，需要一个 C 脑来监视 B 脑，因此这个理论也是容易引起无限退行的。任何一种把意识看成不可分的至高无上的控制器的理论，要么就不可理解，要么就会引起无限退行。因此建立有效的关于意识的理论模型，需要对意识进行功能分析，把复杂的现象分解为简单的功能；另一方面，需要一个能够实现自我表述系统的理论范式，以解决意识内省的问题。

与意识的关系最为密切的两种心理功能是注意与记忆。注意是指进入意识领域的信息所受到的局限，用日常语言来描述这种现象就叫作“心无二用”，人不可能同时注意很多事情。大脑的神经网络处理信息的基本形式是并行处理的，只要设有目标或动作上的冲突，人可以同时作几件事，例如我们可以一边走路一边看书。但是注意的中心在一个瞬间只能集中于一处，如果注意看书，走路就可能撞在树上。因此意识就体现出在并行的大脑中的串行过程。有意识的注意对信息进行选择和限制的现象是人所共知的，但我们更感兴趣的是这种选择和限制的理由。有些人认为这种限制起因于有意识信息处理资源的限制，但是由此又产生一个新的问题，即为什么无意识过程可以并行处理而有意识的信息处理却不能分配更大的资源呢？为了回答这个问题首先要弄清意识到底起什么作用。我们认为，意识的作用就在于对学习进行组织。从进化的过程来看，意识是高等动物进化后期的产物。对于象寄生蜂那样按固定程序动作不需要学习，或是象蜜蜂那样按固定程序学习的动物来说，不需要意识的参与。与其他动物相比，人类大脑之中需要通过学习而建立其功能的部分要多得多，而且具有多重层次，相互关系也很复杂。同时对多层次的大量模块进行学习，势必遇到指数爆炸的障碍。因此只能采取进化中所用的逐层建构

的方式通过长期积累进行学习，意识就起到组织这种逐层学习的作用。为了避免指数爆炸，每次学习的规模必须限制在短桥原理所规定的范围之内，这也就是有意识的注意对信息进行限制的理由。对于按照本能的，或经学习获得的固定程序进行的行为，由于其中不包含学习过程，没有指数爆炸的危险，只要避免了目标或动作模块之间的冲突，大规模并行处理是可行的。

记忆与意识的密切关系是非常明显的，凡是进入意识的信息必然是被记忆的信息，未被记忆的神经活动都不可能是有意识的过程。我们知道自己的有意识的行为就是因为能够回想起这一行为。在梦游状态下人可以作很多事，甚至可以进行简单的对话，但是事后本人完全回忆不起这段梦游的经历，因此我们只能认为梦游是无意识的过程。记忆是学习的必要条件，特别是要建立预测关系时，对于过去发生的事件的记忆是不可缺少的。当然并不是所有的记忆都会被意识到的，在某种技能熟练的过程之中，各种动作操作的控制能力的生成和优化显然引起相应的神经网络模块的持久性变化，因此不能不算是一种记忆，但由于控制信息是下行信息，不可能进入上层的意识，因此我们只可意识到动作或思考的效果反馈，却不可能意识到动作或思考本身。对于思考来说，我们只能意识到思考的一系列结果，而这些结果产生的过程，却是不可能意识到的。我们所能意识到的只能是途经感觉信息处理的上行通道的信息。

关于记忆的一个重要问题是在众多的感觉信息之中怎样选择应该记忆的信息，以及应该以什么形式进行记忆。这个问题的难点在于，从适应的角度看，我们应该记住将来会有用的信息，但是是一种信息在将来是否有用可能是很难预测的。因此记忆的强度（不易忘却的程度）是受多方面因素控制的，首先是取决于对信息的评价绝对值，这种评价一部分来源于本能，例如象疼痛或疲劳这样的保护性或经济性评价，再就是与预测有关的新奇性评价，以

及象自己的姓名这种经学习得到的评价。另一个因素就是实际使用的频度,其根据是过去的使用频度往往能预示未来的使用频度。因此人的记忆不仅仅取决于首次接受该信息时的评价,更取决于其后使用(检索)该信息的频度,这个过程有点象动态存储器(DRAM)的“刷新”过程。为了提高记忆的经济性,遗忘不必要的记忆,人的记忆分为短期、中期、长期的三种不同的记忆期间。脑神经科学的研究结果认为,短期记忆时间为数秒,其机制依靠神经中的电脉冲,中期记忆时间为数分至数日,依靠神经元突触传递能力的化学变化,长期记忆为数月乃至一生,其物质基础可能是神经网络结构的变化。相对短期的记忆反复被检索可以导致该信息进入更长期的记忆,而信息的评价价值较高则会促使其多次被回忆检索。有了这样分期记忆的机制,可以使得实际上有用的信息被记忆得较深,而无用信息则被遗忘,保持记忆的经济和有效。从记忆的形式来看,可以把记忆分为故事型和模型型的两种,短期记忆较倾向于故事型而长期记忆则倾向于更有用的模型型。儿童都喜欢听故事,而且对故事记忆较为完整,当故事积累多了以后,就可以对其他未听过的故事作出某种程度的预测,甚至可以编故事,这就是建立模型的过程。成年人头脑中已经建立了大量的模型,对故事情节的记忆则不再趋于完整,倾向于用某种模型去对故事进行解释和评价,达到“仁者见仁、智者见智”的境地。这里所说的“故事”当然包括自身的经历和对环境的直接观察。了解了“故事”与“模型”的区别,我们就可以理解为什么人可以对某种事物回答他在接受信息时无法预料的问题。如果人的记忆是纯粹“故事型”的,当然不可能作到这一点,但是当故事被加工成模型之后,对模型的操作就可以实现对不同提问的回答,正如我们根据少数视图完成对一个立体的推测之后,可以给出该立体的任何方向的视图或剖面。

当认知主体获得了大量模型之后,就可以区分故事中可预测

的平庸之处和不可预测的精彩之处，从而知道哪些信息是新奇的，对于修改和重建模型是有用的。这种由预测原则产生的新奇性评价对于意识和注意是一个重要的支持，而它本身并不需要意识作为前题。动物行为学家用火鸡作过一个有趣的实验。雌火鸡对雄火鸡的鸣叫会产生反应，实验者用与雄火鸡鸣叫频率相同的声音也可以引起相同的反应，但经过反复刺激之后，反应消失，而用稍有不同的频率刺激时反应又重新出现。对于这种现象往往用对某一频率刺激的神经疲劳机制来解释。但以下的实验推翻了这一看法，即不改变频率只减少音量也能引起反应，这是疲劳所无法说明的。这种学习过程可以用预测来加以说明，当某种现象成为可预测时是不引起注意的，而当预测产生偏差时则会引起注意，这样对音量减小所发生的反应就是很自然的了。这样的评价过程本身是很简单而可以自动完成的，它是意识与注意的必要功能，但它不需要意识那样复杂的解释。

预测偏差引起意识的注意这种功能可以解决大脑这种复杂系统在学习中所遇到的“责任局部化”问题。明斯基在《*The Society of Mind*》一书中把大脑看成是由众多的代理人（Agent）组成的社会，代理人的概念相当于本书的模块。在一个上下级和左右邻关系非常复杂的由大量模块组成的“社会”之中当发生问题时需要追究谁的责任，即修改哪一个模块，是复杂系统学习必需解决的问题。明斯基对此提出了两个解决方案：

① 全局学习方法——按最高目标的成败对所有参与此工作的代理人进行赏罚。

② 局部学习方法——按直接上级的成败对代理人进行赏罚。全局学习方法很难收敛，因此显然是不可取的。局部学习方法也有问题，上级的失败可能是上级组织错误，也可能源于某个下级的失误，对这些不加区分而对所有直接下级进行赏罚显然是不合理的。如果每个代理人都能预测自己行为的结果，当结果与预测

不符时引起意识的注意，启动学习程序重新学习，那么就能够最合理地实现责任局部化，学习只涉及一个模块，修改的方向也是确定的，其结果是减少今后的预测误差。这样就可以避免学习中可能发生的指数爆炸。

注意的信息量限制来源于短期记忆的记忆容量限制。心理学家米勒（Miller）在1956年发表的一篇著名的论文《神秘的数字 7 ± 2 》中概括了他的研究成果：人的短期记忆的容量限于7个项目。通常我们使用的词汇都不超过7个音节，超过7个音节的词汇都是复合词；唐诗每句最长是7个字；人一眼能分辨物体的个数不超过7个；超过7位的电话号码需要分段记忆；如此等等。这个神秘的数字已经为许多人所熟知和承认，但是其理由仍然不为人们所知晓。中野馨（Nakano Kaoru）在和作者闲谈中也提到过一个怪现象，即用联想存储器存储动作串时也很难超过7个动作。虽然我们目前还未能对这个神秘的数字给出严格的数学证明，但是可以想象，这个数字就是短桥原理中短桥的长度，是由于指数爆炸的限制所造成的，而不是进化中的偶然现象。另一方面，一个信息块中的元素过多，对于灵活运用这些信息是不利的，把大块分成若干小块的阶层化组织，有利于各小块重新组合成别的大块，提高学习效率。例如一个武术运动员如果只会套路而不能将动作分解应用，在实战中是没有用的。

只有进入短期记忆的信息才能进入意识，但并不是所有进入短期记忆的信息都进入了意识。短期记忆不只一处，听觉、视觉及其他各种感觉都有各自的不同处理层次的短期记忆，短期记忆中的信息可以是当时的传感信息经过加工处理得到的感觉，也可以是在思考时对中期或长期记忆进行检索或模型操作所得的结果。而这些信息中只有获得当时相对最高的评价绝对值的才能进入意识。评价价值不仅取决于本能决定和在过去学习中获得的重要程度，以及信息新奇性所引起的重要程度，还取决于前置注意的

强度。前置注意是在某信息到达之前由其他信息通过模型操作的预测所决定的将要接收重要信息的预定短期记忆场所。例如在“鸡尾酒会效应”中由自己姓名所引起的其后对同一人物语音的注意。

由以上的分析可以想到一个问题，就是象“感觉”这样的词汇是否被使用得过于笼统了。根据感觉的不同信息处理水平，应该有各自不同的称呼以资区别。低层次的感觉是直接由传感器 (*Senser*) 把物理量变换成统一的神经脉冲或电信号，这个过程可以称之为“传感” (*Sense*)。无论在机器中还是不同复杂程度的生物之中，传感是普遍存在的。大多数的传感并不涉及记忆而直接造成某种反应，如同草履虫或寄生蜂以及一些自动调节系统中常见的那样。但是在日常生活中我们所说的象冷热、疼痛、疲劳、颜色等各种感觉，则更具有主观色彩，通常指的是传感信息经过多重加工处理之后进入意识，或至少是进入短期记忆的带有评价信息的感觉信息。这种主观意义上的感觉的特点是具有评价因素和事后可以回忆起来。对于这种进入意识的感觉可以称之为“感知” (*feel*)。正如诺尔曼所说的“由痛苦的刺激作出应该把手缩回来的决定并不需要什么了不起的认知。但是要在事前回避这种状态，认知就成为必要的了。”因此对于按固定程序动作而不学习或只按固定算法进行学习的动物来说，意识和主观意义上的感觉都是没有必要的，只要有“传感”的功能就足够了。只有具有高度复杂的神经系统，需要进行逐层建构的学习，并在学习中对客观世界建立预测模型的认知主体，才有必要具备意识功能。

把人的心理活动分为有意识和无意识的部分，是弗洛伊德的功绩。弗洛伊德的错误在于把人的下意识的本能看成是天生反社会的。如果人的生物学本能是反社会的，我们就无法解释人们何以会建立起社会。这个问题在基督教文化的深层精神结构之中并不会引起什么矛盾，维持社会存在所需要的道德良心可以从上帝

的教诲中得到，而无需从本能中寻找原因。但是在一个上帝的存在并不天经地义的文化背景下，我们只能从本能之中寻找社会得以产生的原因。动物集群生活的根据是合作所产生的经济利益和适应度的提高。人类的祖先类人猿和现在的近亲黑猩猩都是群居的动物。在从猿到人的进化过程中，引起人类智力高度发展的是智力与文化的协同进化，而群居和规模越来越大的社会显然有利于文化的创造和传播，人的社会性本能在进化中体现出越来越大的适应度贡献。因此弗洛伊德所看到的下意识本能与社会约束之间的冲突应该看成是低层次本能与高层次本能，眼前局部利益与长远全局利益之间的权衡。

如果把意识理解为完全独立的中央控制机构，是藏在大脑中的“小人”，或是机器中的“幽灵”，那么意识要么成为不可分析和不可理解的事物，要么就需要去寻找“小人”背后的“小人”，控制“幽灵”的“幽灵”，结果是陷入无限退行，毫无结果。为了使意识现象成为可理解的，需要从意识与无意识之间的相互关系和相互作用着手进行分析，建立关于意识的“机械论”的基于因果关系的模型。这样就可以不需要“小人”帮助，从机器中赶走“幽灵”。由于意识具有内省的功能，所以人脑是个自我表述系统，建立这种系统的模型，需要采用广义进化范式。按照广义进化论的基本观点，意识是与无意识相对而言的，意识与无意识构成一个软硬结构，意识是其中的软结构，无意识是硬结构。意识在无意识控制下活动，对应着硬结构对软结构的支持；意识活动的结果逐步增强无意识的功能，对应着软结构对硬结构的建构作用。为了便于理解意识与无意识的关系，这里给出一个极度简化的形象化模型。

在图 4.8 中用眼和手代表传感器和执行机构，在眼的感觉过程中包含对动眼肌的控制，在手的动作中又包含了触觉的输入，这样表现是为了强调动作感觉连合和主动感觉的观点。大方框之内

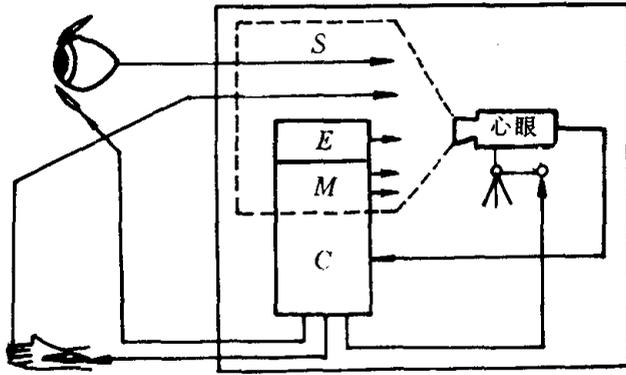


图 4.8

表示大脑内部，安在三脚架上的摄像机(Video Camera)代表“心眼”(mind's eye)，虚线内部是对于“心眼”来说可见的部分，也就是可以被意识到的信息，包括当时来自传感器，经过信息加工的感觉信息(Sense)，或是由模型操作或其他检索方式从记忆(Memory)中得到的信息，以及这些信息所带有的感情(Emotion)评价信息。“心眼”注意何处，是由控制器(Controler)控制的，体现硬结构对软结构的支持；而从“心眼”得到的信息通过某种学习算法对记忆(M)、感情(E)和控制(C)进行修改，体现软结构对硬结构进行建构。控制器(C)位于“心眼”看不到的地方，相当于硬结构即无意识的部分，控制器对“心眼”的控制(在图中表现为由C指向摄像机把手的箭头，表示用把手控制摄像机的方向)，其方法与控制外部执行机构的方法一样，可以通过学习逐步熟练和增加其功能。这个模型主要用来说明意识与无意识之间的软硬结构相互作用，各种功能集中表现只是为了易于理解，在更复杂的用于说明各种心理现象和工程可实现性的模型中，各种功能是分布式存在的，在大脑中当然不存在一个具体的“心眼”，“心眼”的概念只是意识功能的集中描述。

为了更接近大脑中的实际过程，为制造有意识的人工认识主

体提供原型，需要给出意识的分布式模型，把图 4.8 的简化模型表现为分散的神经网络模块形式。可以假定所有与意识有关的较高层的神经网络模块都具有短期记忆的功能。模块具有休止与兴奋的不同工作状态，而兴奋又具有不同的兴奋度，可以用一个连续量（正实数）表达，休止状态时兴奋度为零。模块的兴奋度取决于当时由感觉或思考引起的输入信息在当时环境下的评价绝对值，以及各模块评价之间的相对比较。模块兴奋时伴随短期记忆，如果假定模块在兴奋之后的记忆按相同的时间常数以指数函数衰减的话，那么兴奋度越高，衰减到某个固定水平的时间越长，其短期记忆时间也就越长。当然各种模块的时间常数是很不一样的，一般来说，上层模块比下层模块的时间常数要大，例如关于处理词的模块比处理音节的模块时间常数大，更上层的处理词组、语句、语义、情绪的模块则依次具有更大的时间常数，因此其短期记忆的时间也更长。人在睡眠状态下，高层次模块进入休止状态，而低层次的模块则会由外界刺激诱发或自发地进入兴奋状态，形成梦境，此时上层模块对下层模块的连续性和全局模型的控制不起作用，从外界来的感觉信息的约束也不完全，因此梦境往往是支离破碎而且全局不合理的。由于上层模块不工作，只有在梦醒之前的一部分短时记忆能够在睡醒时很短的时间内回忆起来，通过觉醒状态的回忆而被记住，大部分的梦都是被忘却的。低层次模块在睡眠状态下仍处于工作状态，在必要时使人觉醒，增加了睡眠状态下的警惕性，显然在进化中属于适应的性状。

在图 4.8 中我们用“心眼”的视点（摄像机的指向）来表示注意。在分布式模型之中，注意则表现为各模块之间互兴奋和互抑制的评价竞争结果，评价最高的模块成为注意中心，注意中心的信息再参考环境信息驱动模型操作控制意识流的走向（相当于图 4.8 中的控制器），完成硬结构对软结构的支持。为了使评价竞争只产生单个或少数的模块，在整个的由大量模块组成的神经网络

系统中，与意识有关的具有短期记忆功能的模块应遵循总兴奋量有限的原则，由于总兴奋量的限制，使得高度兴奋只集中于单个或少数模块上。

意识的重要特征是具有内省功能，人可以知道自己正在干什么或想什么。自己看自己有两种方法，一个是用镜子，一个是用照相机。镜子作为模型的优点是其实时性，没有时间滞后，但其不可取之处在于不看镜子时就不知道自己在干什么，而整天盯着镜子又什么也干不成了。照相机或摄像机的模型则没有这个缺点，人们可以干自己的事，事后需要时拿照片或放录像来看，但是总是滞后一点时间，而且需要记忆功能。意识所采取的就是摄像机加录像机的模型，因此严格地说，我们并不知道自己“现在”在干什么、想什么，只知道自己“刚才”在干什么、想什么。而真正的“现在”进行的则是对短期记忆的一种回忆，即我们称之为“追溯”的操作，正是追溯操作，使认知主体具有内省的能力。追溯在录像机模型中相当于倒带重放的操作，但是认知主体的短期记忆与磁带的结构和检索方式有很大区别，短期记忆不象磁带那样只有一条，而是分为不同时间尺度的很多层次，所记忆的内容也分别属于不同的领域，因此追溯时要对这些检索因素作决定，这些检索因素取决于追溯的目的和追溯的技能，追溯的技能与其他技能一样，可以在实行的过程中根据结果和评价进行修正和提高。追溯所得的信息就是意识内省的结果，属于软结构，而追溯所需的操作技能属于硬结构，通过追溯操作获得结果是硬结构对软结构的支持，而追溯结果通过评价对各种技能的模型包括追溯技能本身进行修正和再建则是软结构对硬结构的建构。

不同层次、不同时间尺度的短期记忆对硬结构中的模型进行修正和再建是一个复杂的过程。在外界信息进入短期记忆即被意识到时，首先经过了评价的过滤，使得只有对修正模型有意义的重要信息才进入短期记忆，而已经被智能模型完全预测的信息则

不会得到较高的评价。信息的意义不仅包含其重要性，而且包含其对哪些方面重要的目的信息，在大多数主动感知的场合，获取信息的目的是明确的，而在突然的信息中断过程中，引起中断的评价或中断之后对信息的解释都要依靠某个智能模型，而这些智能模型也都含有一定的目的信息作为其起动检索因素。进入短期记忆的信息可以根据其附加的目的信息进行整理，做到“物以类聚”，这样即使在大时间尺度的短期记忆中经常出现很多中断，不同的短期记忆中的“故事”也不会相互混淆，例如每天晚上播出一集的电视剧，尽管观众在两集之间进行了一天的日常活动，也不会把日常生活的事情混入电视剧的情节中去。

“故事”是生动的智能模型，而模型则是经过提炼的“故事”。故事是通过记忆直接得到的，比模型带有更丰富的信息，或者说包含很多对于模型来说多余的信息。故事本身可以作为模型来用，具有预测能力，因此也可以说故事是原始的智能模型。在接受新的故事的过程中，老故事或其片断被检索用来对新故事进行预测，如果预测成功，新老故事之间共同的特征被加强；老故事的检索键信息也被加强，如果预测失败则相应的信息被减弱，而新故事的不可预测部分作为新的原始模型被记忆。老故事在这样反复被使用的过程中被检索的特征和各故事共同的特征被加强，而其他多余信息则被忘却，逐步被精炼成能在适当的时机被调用，而且预测能力很强的模型，此时老故事本身大多已被忘却了，而模型则固化到长期记忆之中。

以学习汽车驾驶为例，初次坐在司机位置上的人，会注意方向盘、手柄、踏板、仪表之类，通过一个阶段的学习，逐步在头脑里建立起有关的智能模型，可以正确预测汽车本身对操作的反应，如方向盘转动多少可以拐多大角度，刹车踩多大的力可以获得多大的负加速度，油门踩多少可以得到多少正加速度，等等，此时驾驶员不再意识到方向盘和踏板，只想着加速、减速、拐弯即

可。在进一步的上路训练之后，驾驶员的头脑中建立起有关交通规则模型，此时交通规则的一些条文可能已不记得，但可以对交通信号和各种标识作出正确反应，此时驾驶员已不再意识到交通规则，而只意识到要去何处和走哪条路，如果经常走同一条路，对于在何处拐弯已经建立起模型，驾驶员只需要注意突发意外情况即可，在正常情况下完全可以靠下意识的模型自动驾驶，可以边驾驶边聊天或想别的事情，在必要时值得注意的信息会通过正确的评价系统产生中断，自动引起驾驶员的注意，例如当刹车失灵时，踩刹车时车子的减速情况与驾驶员头脑中模型预测不一致，立刻会引起驾驶员的注意，采取必要的对策。学习驾驶的全过程体现了软结构（意识）对硬结构（无意识领域的智能模型）进行逐层建构，硬结构的生长对意识进行支持的过程。只有当下层不再需要意识的关照逐渐走向自动化时，意识的注意才能集中于上一个层次。

象这样通过意识与无意识的软硬结构相互作用逐层建构起来的智能模型，就是诺尔曼所说的信念系统。智能模型存在于无意识的领域，不能由意识直接观察，但它决定意识的活动方向，同时又会由意识活动的结果而发生变化。从宏观上看，信念系统取决于遗传特性、文化环境和个人经历，但是由于其软硬结构相互作用的复杂建构过程以及其中的随机因素，除了一部分收敛因素之外，信念系统是不可预测的，这也就造成了人的行为的某种程度的不可预测性。这里所说的收敛因素是指某些在强进化压力下保持不变的因索，例如鲨鱼和海豚虽然经历了完全不同的进化途径，但是由于在水中的高速运动，其体型呈相似的流线形；又如虽然人的遗传特性和经历千变万化，但在同一个社会环境下生存时会有大致一致的基本道德观。由于东西方文化背景的区别，中国的一些学者不太喜欢“信念”（*belief*）这个词，由于它具有“宗教信仰”的含义，作为科学的名词不太适宜，我们不能象西方人

那样假定所有人都具有某种宗教信仰，因此我国学者戴汝为教授认为在这种场合应该把 belief 译成“感受”更为确切。这里的“感受”指的是一些“只可意会，不可言传”的东西，例如可以通过语言传授几何定理，但不能说出如何理解和使用这些定理，只能用例示间接表达，几何证明的能力要在作习题的过程中逐步培养。这种不可言传的东西就是认知主体中的智能模型，而我们只能意识到模型操作的结果，而不能意识到操作过程本身，因而也就不可言传了，如同骑自行车的技能一样。

自由意志，或曰意志的自由是与意识密切相关的一个重要概念。意志自由的概念在东方文化背景之中并不占什么重要的位置，现代的中国学者也不太热衷于讨论这个问题。但是对于西方文化背景来说，意志的自由却是极为重要的基本出发点，是建立多种理论的必要前题假说。因此当我们对意识给出机械论的模型时，对于自由意志问题也需要作出与之相一致的解释。明斯基早已意识到关于大脑或精神世界的机械论解释与自由意志之间的难以调合的矛盾。如果人脑是一个用自然科学的眼光可以理解的系统，那么这个系统的运行就得依据确定的自然规律即决定论的原因，或是概率论的偶然。无论是受因果律还是偶然因素的操纵，在这个系统之中都没有自由意志或灵魂存在的余地，因此明斯基把意志的自由称之为除了必然和偶然之外的“第三条道路的虚构”。在远古时代意志的领域是相当广阔的，日月星辰都是一些神仙，风云雷电、草木动物都有灵魂，但是若干世纪以来随着科学的昌盛和泛灵论的退潮，意志的领域成为一个不断衰亡的帝国。对此我们是否可以完全采取现代科学的观点，把意志的自由看成过时的神话或虚构而排除掉呢？明斯基认为这是作不到的。因为我们的意识形态和所作所为在极大程度上依赖于关于意志的神话。例如社会生活的规范在很大程度上依赖于“责任”的概念，而“责任”的概念如果失去个人的行为可以由自己的意志自由地加以决定这一

前题，几乎没有任何意义。如果个人的行为是被某种自然规律或掷骰子所决定的，那么这样的行为就没有理由受到任何赞扬或责备。在这种情况下，我们将失去罪行与德行的区别，也无法对孩子们进行区分善恶的教育。

按照明斯基的看法，要么牺牲科学的原则，要么牺牲社会生活的规范，要想同时保全二者就必须接受物理世界与心理世界相互割裂的局面，放弃理论统一的理想，而这三种选择都是令人不快的。然而这种矛盾状态并不是心理学与物理学之间必然存在的，它来源于基督教文化的理论体系中自由意志所处的核心地位。笔者曾经与一位基督教的传教士探讨过这样的问题：“你们认为世上万事万物都是由上帝创造的，那么上帝为什么要创造出魔鬼与自己作对呢？”那位传教士是这样回答的：“上帝并没有创造魔鬼，只创造了天使，上帝给了他们自由意志，有的天使由自己的自由意志变成了魔鬼而反抗上帝。”从这个说法中可以看出自由意志在理论上的重要地位，它是一种连全能的上帝也不能控制的力量。亚当和夏娃被逐出伊甸园的故事也是一个相似的过程，人类偷吃了苹果而获得了区分善恶的能力，从而能按自己的自由意志选择做什么和不做什么，由此才产生了对于罪恶进行末日审判的需要。

如果我们离开基督教的文化背景，自由意志的问题就不会引起什么矛盾。我们完全可以接受物理学意义上的必然和偶然，同时在这个基础上建立与自由意志有关的具有一致性的理论，而且与社会生活规范并不发生冲突。只要不把自由意志当作形而上学的前题看待，我们就可以仍遵循科学的原则，认为头脑中只存在有规律的必然和随机的现象，而不需要虚构出一个既非因果又非偶然的“第三条道路”。我们把意识看成是注意与记忆结合产生的现象，而自由意志作为一种现象就是可被意识到的决策过程。我们认为自己持有意志的自由的根据就是我们能够意识到自己在决

策过程中所进行的选择。不具有意识功能的动物也有各种决策过程，这种决策由各司其职的神经模块之间的竞争来完成，并不需要意识的参与。人的决策过程中大多数也是在下意识领域完成的，只有在未知因素较多或评价难分胜负时决策过程才会引起意识的注意。决策过程之所以需要被意识到而进入记忆有两个原因，第一个原因是人的价值观体系比较复杂，而且有不少价值观要通过后天的学习建立起来，并在学习中加以修正；第二个原因是人能够进行大时间尺度的有预谋有计划的行为。对于价值观只在进化过程中改变而在个体一生中只取决于基因不进行学习的动物来说，记忆决策过程是没有意义的。ES 学习的结果也可以看成是价值观的变化，但对于多数动物来说，由于不具备中断恢复的能力，不可能进行大时间尺度行为，在每个行为中模块可以自始至终处于兴奋状态，无需特别的记忆就可以决定怎样对价值进行修正。但是由于人能够进行大时间尺度的行为，其间往往受到多次中断，行为决策与行为结果之间时间间隔很长，为了根据行为结果对形成决策的价值观进行修正，必需使决策过程进入意识，从而和目标信息一起存入记忆之中，使得在行为结果出现时可以根据目标信息检索出相应的决策过程记忆，以决定如何对决策所参照的价值观进行修正。在遇到由于决策的恰当或不当而引起行为的成功或失败时，人们往往会体会到庆幸或后悔的情绪，庆幸与后悔一方面体现正的或负的评价，一方面伴随着对决策过程的回忆，这就为价值观的修正提供了可能性，在众多的个体组成社会时，全局利益与个体利益之间，眼前利益与长远利益之间有时会发生类似于囚犯困境那样的冲突。为了防止局部和眼前利益破坏全局的长远利益，需要每个社会成员建立起与社会相适应的价值观，而当某些社会成员的价值观偏离全局最佳点时，需要有社会的赏罚作为个体学习系统进行修正的依据。为了使修正和学习正确进行，准确的责任局部化是必不可少的。当决策出现错误时，应该修正的

是造成这一错误的价值观，由此就足以支持责任的概念，而无需意志的自由作为形而上学的基础。意志的自由仅仅是整个学习系统之中处于中间变量位置的概念。综上所述，一个古典的机械论的系统不论是缘于因果律还是随机性，都不会涉及责任和赏罚的问题，但是对于可学习的系统来说，责任与赏罚的概念就是必不可少的了，无论这个系统是否具有意志的自由。

4.8 创造与灵感

生物的进化是自然界中最活跃的创造过程。为了理解人脑的创造能力，生物的进化是最好的理论范式。图 4.9 表现了生物进化与文化进化中创造过程的对应关系。图中①的对应关系

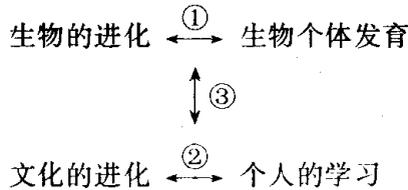


图 4.9

在很早以前就被生物学界所熟知，生物的进化过程与生物个体发育过程之间存在着不十分严密但很明显的对应关系，即胚胎发育过程在某种程度上重复着从单细胞生物到多细胞生物的进化过程，因此胚胎学曾经成为进化论的重要证据。从广义进化论的观点来看，这种重复可以用硬化机制来解释。在基因的变异和生物成体性状变化之间存在着漫长的因果链，即胚胎发育过程。在变异受到成体性状是否有利的考验之前就已经受到胚胎发育能否完成的考验。甚至可以说，与搜索生物成体适应环境的性状相比，搜索成功的制造工序是进化的更为核心的问题。由于后工序对前工序的硬化以及互适应协同进化的不可逆性，前段工序被顽强地保存下来，可以在几十亿年中不发生大的变化。换句话说，除了重

重复进化过程，生物无法知道其他的制造自己的方法。

图 4.9 中②所示的对应关系不象①那样明显。一般看来，个人学习过程并不一定重复文化发展史，例如人不需要先学古文再学现代文。在这里我们需要区别看待两种不同的文化，即进化型文化与非进化型文化。进化型文化的典型例子是科学，非进化型文化的典型例子是宗教。进化型文化的最主要特征是具备软硬结构，因此发展很快而且水平不断提高。科学中有争议的部分是软结构，处于已知与未知之间的领域，通过研究达到众说归一的阶段，则该领域硬化，软结构被推向原来未知的领域。按这种方式发展起来的文化具有较强的阶层性，不具备前层知识则无法理解后层。在科学发展史上有不少多人同时创造类似理论的事例，这说明在科学发展的当时水平已经具备了创造该理论的必要条件。当这些条件不具备时，不仅不可能创造出这一理论，甚至理解它也是很困难的。对于进化型文化来说，图 4.9 中②的对应一目了然，只要看一看标准的教科书，就可以看出关于科学的部分的学习顺序与科学发展史的顺序是大致一致的。

图 4.9 中的③所表示的对应关系是很有启发意义的，它是①和②两个对应关系之间的对应关系，表现了文化进化与个人学习之间的关系和生物进化与个体发育之间关系的某种相似性。这种对比可以使我们对于创造过程有一个更深层的知识。我们通常所说的创造是指通过头脑的活动产生出前所未有的思想、理论、方法、技术、作品等，加入到文化总体之中去，从宏观上看，是文化进化的软结构对硬结构的建构过程，而这个建构又是受到硬结构支持的。通过对应关系③，可以看到文化进化中的创造与生物进化中的“发明”相对应，而文化进化中的创造与个人学习中的理解之间的关系则类似于生物进化中的“发明”与个体发育中基因控制的建造之间的关系。从这个对应关系出发可以看到，象生物个体发育重复进化过程一样，个人学习的理解也是重复文化进

化的创造过程。所以说，理解就是再创造。理解与创造的共同点是，两者需要相同的前置知识，相同的认知模型，同样受到短桥原理的限制；二者的区别则在于，理解具有明确的目标，通常不需要搜索过程，而创造则没有明确目标，往往需要困难的搜索过程。由此可见，不知道创造的机制，就不可能知道理解的机制，只是理解由于免去了搜索过程，比创造的效率高得多，几千年文化进化所创造的知识体系，可以在几十年或十几年中被理解，就象生物个体几十天，几个月的发育，由于没有搜索过程，可以重复几十亿年的进化。

从文化进化与生物进化的类似性对比之中可以得到的另一个重要的启发是，在文化进化的软结构之中所发生的创造过程与生物进化的软基因中的“发明”过程可以用相同的原理来实现，这个原理就是变异与选择的原理。当然，人脑之中的创造过程与生物的进化过程既有相似之处，也有很多不同之处。生物进化中的变异的机制主要是基因的突变和重组，而选择的标准则是达尔文适应度，即包含了发育成长全过程在内的综合繁殖效率，评价准则是比较单一的；而人脑的创造过程中的变异机制与选择时的评价准则则要复杂得多，但是按照变异与选择这个思路可以对创造过程建立起比较有现实意义的模型。创造过程中的变异可以有多种不同的方式，最简单的是扩大智能模型的应用范围。在通常的行为和思考过程中，每一个智能模型都有其特定的应用范围，从而可以在解决特定问题时能够根据问题的特征直接调用智能模型进行预测。但是当认知主体遇到新问题时，会落入不知该调用哪个智能模型的处境，此时如果问题是处在短桥距离之内，认知主体之中有可能存在着可用的智能模型，只是缺乏调用这一模型所须的索引信息，这就需要动用搜索的方法，对于合用的可能性较高的智能模型进行逐个的调用试验，调用如果成功，调用时所用的检索信息就被固定下来，使得以后遇到同类问题时可以直接调

用这一模型。变异的另一种较为复杂的方法是智能模型的重新组合，有点类似于遗传的杂交过程，在思考过程中由问题与环境信息调用不同的智能模型，不断编出新的“故事”，当这一故事较好地符合当时需要的评价准则时，就会进入意识而被记忆，以这个“故事”为雏形，逐步加工成新的智能模型。无论用什么方法，变异的的目的都是产生大量的新的智能模型，而选择的任务则是从中挑出对解决问题有用的部分。因此选择时的评价准则就不会象达尔文适应度那么单纯，而是随要解决的问题而变化的。

灵感是与创造密切相关的思维现象，结合探讨灵感产生的机制可以对人脑的创造过程作出更详细的解释。灵感作为一种思维现象早已引起很多研究者的关注，我们可以引用刘奎林在《灵感发生论新探》（载《关于思维科学》钱学森主编，上海人民出版社，1986）中所述的王国维对于创造境域所划分的三个阶段作为灵感现象的概括。第一阶段是“昨夜西风凋碧树”、“独上高楼，望尽天涯路”；第二阶段是“衣带渐宽终不悔，为伊消得人憔悴”；第三阶段是“众里寻他千百度，蓦然回首，那人却在，灯火阑珊处”。这里的第一阶段是指进入问题环境，第二阶段是对问题百思而不得其解，第三阶段则是一种顿悟，此时会有一种茅塞顿开，豁然开朗的感觉。

上述关于灵感的一些现象只是我们的意识可以觉察的部分，单凭这些现象还不足以推测出灵感产生的机制。为了对灵感的现象作出解释，我们还需要知道无意识领域中的信息处理过程。灵感现象是意识与无意识相互作用的结果，这一点已经是很多研究者的共识，但是对于无意识领域中到底是什么样的过程，则是众说不一，缺乏有说服力的看法。大多数的看法都过于模糊，不具备工程可实现性。其中比较清晰的一种说法是潜意识推论说，基本思想是在意识领域提出问题之后，无意识的领域为之进行一系列的推论，当推论得到合乎要求的结果时，再进入意识，表现为

灵感的产生。由于用计算机进行推论的问题在人工智能领域已经进行过广泛的研究，所以潜意识推论说的工程可实现性是显而易见的。但是这种理论也还有不少缺陷。首先是推论的方法问题，据我们所知，并不存在一种“发明的逻辑”，因此通过什么样的推论可以最终产生灵感的效果是一个未解决的问题。其次是复杂度问题，如果推论是一个搜索过程，那么应该如何避免指数爆炸就是很关键的问题了。第三是效率问题，由于神经元的计算速度比计算机低好几个数量级，即使把并列计算考虑在内，有些逻辑深度较大的问题仍然难以避免串行过程，这就造成对于一些出现很快的灵感在时间上与推论难以吻合。第四是与外界刺激的关系问题，有时灵感会由某种外界因素触发，这是潜意识推论所难以解释的。因此对于灵感的无意识过程还需要进行更进一步的探讨。

为了回避指数爆炸，创造的灵感所面对的问题必须满足一个前题条件，也就是将要作出的创造产物必须处于短桥原理所规定的距离之内。从另一个角度看，这意味着要想创造出一种新的理论，其必要条件是与之有关的前置理论已经存在。在历史上不同的人相近的时间各自独立的创造相同理论的现象是不乏其例的，其原因就是当时的科学水平已经发展到使该理论处于短桥距离之内，或者说该理论所必要的前置理论已经完备了。例如牛顿与莱布尼兹基本上同时发明了微积分，其必要前置理论之一是笛卡尔的解析几何；达尔文与华莱士先后创造了进化论，当时的科学界已经有了拉马克的原始进化学说，而两个人在进化机制上的发明又都是受到了马尔萨斯人口论的启发，其区别仅仅在于达尔文是在偶然翻阅人口论时受到启发，而华莱士则是在思考过程中回忆起马尔萨斯人口论而受到了启发。产生创造灵感的另一个必要条件当然是创造的愿望。这种愿望有时产生于现实出现的问题的需要，更多地是产生于对于现存理论的缺陷的不满，因此即使新理论推翻了旧理论，旧理论仍然可以是新理论的必要前置理论。

例如我们可以说相对论的前置理论是牛顿力学、麦克斯韦电磁场理论和洛伦兹变换。虽然爱因斯坦的相对时空观与牛顿的绝对时空观完全对立，但仍然可以说相对论完成了牛顿力学。从文化进化的宏观角度上看，前置理论的存在体现了硬结构对软结构的支持，而新理论被广泛接受则是软结构对硬结构的建构。

创造灵感的三个阶段之中，第一阶段是明确问题的阶段。在这个阶段中可以意识到的是新问题导致对于新的理论或方法的追求，或是对于旧理论的缺陷所产生的不满。而在无意识之中，所进行的是对价值观的设定，虽然在这个阶段认知主体还不知道新理论或新方法是什么样的，但是新理论应该满足的要求却逐步明确起来了，也就是说为选择机制准备了明确的评价准则。发明创造的灵感产生的机制与回忆起老同学姓名的机制基本上是相同的，第一阶段的作用是在不知道所要回忆的姓名的情况下把要回忆的人确定下来，其根据则是关于此人的其他信息，例如相貌和一些特定的行为。而对于所要发明创造的新理论和新方法来说，在这一阶段就明确了这个理论应该满足哪些要求，对于旧理论的不满，也意味着对于不满之处的相反方向设定较高的价值。

灵感的第二个阶段是搜索阶段。在这个阶段中我们能够意识到的是集中注意力对该问题进行思考，不时有一些新的想法浮现出来，但由于不完全满足要求而被否定掉。在无意识之中进行的工作则是搜索过程，或者说是变异选择过程。从提高效率的经济原则出发，完全随机的搜索显然是不合理的，搜索应该按照可能性大小顺序进行，就象我们平常找东西一样，总是从找到这个东西可能性最大的地方开始，这是与基因突变不同的地方，而和带有性选择的基因重组有些类似。这种有序的变异可以用以下的机制实现。我们假定由问题检索智能模型时依靠某种特征向量，而在检索过程之中总是找出特征相似度最大的智能模型，如果检索得到的模型操作后不能得到满意的结果，则按照 ES 学习准则修

改特征向量，使其与该模型相似度下降，这样原来处于第二位的模型就会上升到第一位，在接之而来的检索中被选择，依此类推，就可以实现有序的变异，直到模型操作结果满足评价准则时唤起意识的注意，进入第三阶段。如果在第一次检索中就得到了足以解决问题的模型，那么这只是一种顺理成章的思考过程，不会引起产生灵感的感觉，而只是已有知识的一次应用，这说明已有的智能模型及其检索信息对这个问题来说已经合乎要求，不需要记忆和变更，因此也不会意识中造成太大印象。反之，随着检索失败次数的增加，在意识中有一种百思而不得其解的经历，而在无意识领域中，这个过程使得能解决问题的方法的评价价值大大增加。评价值的增加有两方面的理由，第一是多次的检索失败意味着智能模型及其检索信息存在着不适应的一面，有必要进行修改；第二是经多次检索失败而仍不放弃这个问题，意味着该问题是很重要的。因此百思而不得其解的过程对于产生灵感是必要的，它造就了很高的评价价值，使得在一旦得到解答时能够造成很高的评价情绪，在意识中造成豁然开朗的快感。长时间思考及多次检索的作用不仅仅是提高了正确答案的评价价值，在思考中得到的部分正确答案也使问题的周围情况变得清晰；而使问题需要解决的核心更为突出，就象在回忆人的姓名的过程中，即使姓名没想起来，与此人有关的信息也会逐步增加，为成功的检索增加可能性。很高的评价价值和丰富的相关信息在冥思苦想中断或结束之后也能保持相当长的时间，这样在我们并未有意识思考该问题时，只要注意力不被更重要的事所吸引，处于散漫的心情时，该问题也会由于其高价值而自动占领检索过程，在无意识中被搜索，当搜索得到所要答案时通过高的评价价值唤起意识的注意。也可能偶然在观察或思考其他问题时碰到有用的模型，引起该问题解决的灵感出现，这些也都依赖于过去不成功的思考所积累的周围信息与评价价值。因此人们常说，机遇只偏爱有准备的头脑。

灵感的第三阶段是豁然开朗的阶段，通常称之为灵感的也多是这个阶段，但是没有前两个阶段这个阶段是不可能到达的。有意识的思考所造成的评价增高和周围信息的丰富，使得在有意识或无意识的检索过程中，或受到外部信息启发所获得的问题的答案获得很高的评价，从而引起意识的注意，并由于其高值的正评价产生很大的快感，造成了豁然开朗的感觉，阿基米德也就是在这种情绪下光着身子从浴池里跳出来，大喊大叫的。当然，灵感出现时只带来了解决问题方案的雏形，要将其变成可以对文化进行建构的成品则还须要乘胜追击作进一步的加工和完善化，但这个过程通常是通过常规的思考过程完成的。

从灵感产生的全过程特别是第二阶段来看，创造过程包含很多的随机因素，具有偶然性。有序的搜索虽然比随机搜索效率高，但是真正具有创造性的模型正是意味着它不处于有序搜索的前几名之内，而对于检索类似度较低的智能模型来说，必定是数量大而且类似度大小顺序并无多大意义，此时的搜索效率与随机搜索没有太大差异。因此创造活动都是风险很大的，失败的结局比成功的结局多得多。当然从统计意义上讲，已有的智能模型越丰富，搜索时间越长，获得灵感的可能性越大。但是搜索不可能无限的进行下去，认知主体只能根据问题的重要性来决定继续思考下去是否值得，因此对问题的评价也是影响创造灵感的重要因素之一。当然，失败的检索并不一定是百分之百失败的，往往可以得到部分解决问题的成果，这种成果能够突出问题最困难的核心，也会增加该思考过程的价值，因此在思考过程中可以得到是否应该继续思考的决策所需的信息，这也是认知主体判断问题是否处于短桥距离之内的根据。

基于上述的对于创造过程的认识，以及非本体论的认识论的立场，我们可以进一步探讨与休谟悖论有关的问题，即归纳法是不是合乎理性的。归纳问题可以从两个角度来看，一个是逻辑学

的角度，一个是心理学的角度，休谟和波普尔都对两者作了严密的区分，而且都是只从逻辑的角度讨论问题，而把心理学的角度放在自己的视野之外。从逻辑的角度看，归纳是不合乎逻辑的，从过去每天太阳都升起的事实不能合乎逻辑地推论出明天太阳还会升起。但是从心理学的角度来看，我们每个人都对于太阳明天还会升起这一点深信不疑。在本书的第二章我们已经通过取消真理这个概念从逻辑上消除了休谟悖论的起因，我们已经知道，无论是归纳还是演绎都不可能证明真理，逻辑的作用在于维护理论的一致性，从而使得理论具有稳定性和预见力。由于本书的目的在于研究认知过程，没有理由象休谟或波普尔那样把心理学的问题拒之门外，因此我们需要解释心理学意义上的休谟问题，即归纳法为什么看起来是可行的，为什么大量符合理论的事实会使我们增加对于理论的信心，为什么掌握大量的事实有利于我们创建较好的理论。

按照我们对创造灵感的内部机制的解释，所谓的“归纳”过程实际上可以分解为一系列的假说验证过程，即对于不同智能模型或模型的组合所进行的搜索过程。虽然事实不可能直接导致新理论的建立，即真正的“归纳法”是不存在的，但是在对智能模型的搜索过程之中，大量的事实却可以起到重要的作用。首先事实可以为智能模型的检索提供索引信息，而大量不同事实所具有的共同特征可以使索引信息更为可靠；另一方面，大量的事实为检索之后的评价提供了更为严格的约束，满足一两个事实的智能模型很容易找到，但由此而产生的理论稳定性差，而且预见力也未受到检验，很可能不满足更多的事实，而且容易出现若干理论难分优劣的局面，而大量的事实和现象则可以克服上述的缺点。公鸡哲学与守株待兔从逻辑上看只是一步与百步之差，但从认知过程来看则有很大的不同。从一只兔子在树上碰死可以引发出很多理论，也可能是纯属偶然，也可能象农夫所期待的那样今后每天

都有一只兔子碰死，也可能是每两天、三天、一周、一个月或一年周期性的碰死，也可能以某种概率随机地碰死，但只凭一只兔子碰死的事实无法区别这些理论的优劣，也就不可能选定一种理论进行预测。因此对于建立理论来说，事实太少是不行的。公鸡哲学的情况就大不相同了，如果公鸡在第10天进行思考，得出每天能吃米的结论，那么这个理论在其后的90天都能进行成功的预测，而10天的经验已经足以排斥两天或三天吃一次米的理论。从逻辑上讲，对于任何有限个事实都可以有无限个理论与之符合，因此一个事实与一百个事实没什么不同。但是从认知过程来看，认知主体中所已经具有的智能模型只能是有限个，其组合也是有限的，一百个事实可以比一个事实造成更强的约束，有助于对理论的选择。

因此，知道大量的事实不仅有利于对智能模型的检索，也有利于对新理论的稳定性和预见力的检验。特别是用这些事实中的一部分作索引而另一部分作为检验时，所得到的理论更为可靠，正如我们过有限个点可以画无限条曲线，但过其中的一部分点所画的曲线正巧使得其他点也都在曲线上则要困难得多。这也就是归纳法的合理之处。特别是当某个理论能够预言未知的事实而又经受住预见正确的检验，则更可以提高理论的价值，因为这足以证明我们在创造理论时并没有参考这些事实，对未知事实的预见是货真价实的预见。当然，过去预测的成功从逻辑上不能保证未来预测的成功，但是在实际的认知过程中我们只能期待这种成功，寄希望于客观世界的相对稳定性，正是相对稳定性假定使得理论的预见力也保持相对稳定，没有这个假定，一切进化与学习在逻辑上都是不可能的。然而，我们也不应忘记这种稳定性是相对的，繁荣一时的物种也可能灭绝，鼎盛的文化也可能衰亡。因此在认知过程之中，智能模型总是随时对自己所作的预测进行着检验，在预测出现偏差时进行修正，乃至建立新的模型。认知的可靠性存

在于这个动态过程之中，而无需任何与真理有关的假说。

4.9 模式识别

通常，模式识别是指把感觉输入的模式信息变换成符号信息的过程。更深入的看，模式识别是一个归类过程，对某一个模式判断它属于已经分好的若干类别中的某一类，如果给每一类加上一个符号作为其名称，那么就可以完成从模式到符号的变换。但是这个符号只对认知系统的外部观察者是有意义的，如果我们不把认知主体看成一个符号处理系统的话，那么分类的结果只要在其后的信息处理中起作用，它被赋与什么符号或是否赋与符号对于认知主体来说是没有意义的。站在符号主义的立场上看，模式识别仅仅是对输入信息的前处理，并不处于与智能有关的核心地位，但是从认知的角度看，归类的基础是分类，而分类是认知的极为重要的核心问题，或者说，识别 (*Recognition*) 就是再认知 (*Re-Cognition*)。

对于把模式信息变换为符号信息的工程模式识别问题，主要可以分为特征抽取和归类判别两个步骤，首先从模式中计算出若干特征的值，然后在特征空间中对特征向量进行归类。但是认知主体中的模式处理还要复杂得多，包括识别目的和背景信息等因素对模式处理的影响，以及通过某种行为主动获得信息的过程，而且归类判别的标准也不是先验地给定的，而是在学习中逐步完成的。这意味着归类即识别的能力是在认知过程中通过学习建立分类而逐步获得、不断完善的，或者说识别与认知是浑然一体的。因此归类所用的判别函数不仅随目的和环境而变化，而且随学习进程而改变。例如我们可以学会根据高度、平整度、支持能力等特征来识别椅子，但在特殊环境下出于别的目的，则又可以根据重量、可抓握等特征把同一个物体看成武器，能够抡起椅子打人。

因为我们把模式而不是符号看成是认知过程中信息处理的基本信息单元，所以模式识别问题不仅由于与认知的直接关系而具有核心地位，而且由于与所有的信息处理有关而处于基础位置。也就是说模式处理在认知过程中具有无处不在的普遍性，不仅对于输入的模式信息需要抽取特征或进行判别，模式信息的联想记忆的存取也需要特征进行索引，智能模型及感觉动作联合也是由模式联想串构成的，智能模型的检索也是靠索引模式的相似性即归类处理，由智能模型的预测所得到的信息依然是模式信息，输入模式经特征抽取所得到的若干特征组成的特征向量仍然是一个模式，评价向量也可以看成是一个模式，因此我们所讨论过的认知系统的所有低层次和高层次的信息处理最终都是建立在模式处理这一基础之上的。反之，我们不用符号而用模式作为认知过程的基本信息单元，也正是由于符号难以成为上述各种处理的基础，以符号为基础的系统不存在内在的距离定义，因而难以实现渐近的学习过程，难以自动具备类似度，也难以自发的形成语义，难以自行实现分类。

把模式识别的整个处理过程分为特征抽取和归类判别两个步骤，可以根据短桥原理来进行划分，其原则就是把判别函数限制在线性范围之内，也就是说特征抽取应使产生的特征空间中的归类是线性可分离的，从神经网络的角度看就是归类可以在单层神经网络中实现，这样就可以保证判别函数的学习过程保持高效率而不会遇到指数爆炸。这样通过学习达到适应分类的认知过程也就不困难了。在工程模式识别系统之中，这种学习可以是有监督或无监督的，有监督时模式的分类结果已知，即分类是由系统外部决定的，所谓无监督学习也是由外部输入模式及特征抽取方法暗中决定学习结果，学习所依据的原则是不明确的，无法保证学习结果的效能。这两种学习方法对于认知系统来说都是不适宜的。认知系统不可能完全指望外部给定的分类，而应该能在系统内部自

行产生概念，所以有监督学习是不可取的。根据丑小鸭定理和认识论的逆镜模型，象无监督学习那样以纯客观标准决定分类的企图也是注定行不通的。指引分类学习即概念生成的只能是认知系统内部的评价准则，所有评价准则归根结底要服从达尔文适应度，但适应度并不适合作为学习的评价准则，学习所需要的是符合适应度即存在原则而又可以直接起作用的原则。从感觉动作联合的角度可以得到一个比较直接的原则，那就是感觉信息的语义是由与之关联的动作决定的，分类的直接目的是引起某种动作或智能模型操作（内化的动作），因此动作或模型操作的成败可以作为最直接的信息用于判断分类是否合乎要求，通过 *ES* 学习或其他学习算法对分类进行修正，对特征进行选择。有效的分类应该满足的一个原则是使归类所引起的动作或智能模型操作获得成功，这也是概念形成的原则。

与判别归类相比，特征抽取要复杂得多，对于模式来说特征抽取的方法有无穷多种，因此在原理上不可能存在客观上最佳的特征，较低层次的特征抽取方法是生物进化史上诸多发明的产物，与信息源的物理或化学基础有关，因此决定特征抽取的算法也是比较复杂的。在较高的层次上，可以根据分类学习的结果决定各特征的重要程度对特征进行选择；较低层次的模式信息处理则主要参考信息源和感觉器官的物理、化学特性和模式信息的性质；对于中间层次来说，则只能依靠普遍的评价准则即特征的稳定性和可预测性来进行选择。

从整个模式识别系统来看，模式的认知与识别也是一个广义进化系统，特征抽取是其硬结构，判别归类是其软结构，特征抽取对判别的支持是显而易见的；判别归类在各种价值观的引导下进行学习，形成有效的分类（聚类），完成一步认知过程，形成概念；与此同时，分类也构成一个层次的特征抽取，分类结果对于更高层次的分类来说是一组特征，这也就是软结构对硬结构的逐

层建构。对于动物或人这样的天然认知主体来说，较低层次的特征抽取能力（算法）是在进化过程中获得的，对于个体来说则是先天的，而较高层次的特征抽取能力则可以靠后天学习逐层建构。后天学习对于低层次的特征抽取也并非完全没有影响，例如猫如果在出生后只看到过纵向条纹而未看到过横向条纹，其神经系统中对横向条纹起反应的神经元就会减少，而纵向细胞会增加，但这仅仅是一种选择过程，象神经达尔文主义所主张的那样，而这两种细胞或模块作为某一阶层的特征抽取算法，则是遗传决定的。由于学习所固有的低效率，从白纸开始的学习是不可取的。对于人工认知主体来说，模拟整个进化过程来产生特征抽取算法是极为困难的，比较现实的方法则是通过研究天然认知主体来得到设计低层次特征抽取算法的原型，或者从文化积累中利用已知的数学方法。而对于特征的选择则可以靠学习来进行，并通过学习过程实现更高层次特征抽取能力的建构。总的来说，设计人工认知主体时还有很多需要预置的信息，例如模块资源分配、学习算法、学习的大时间尺度程序、评价准则的安排等等，人类智能中由遗传决定的部分可能比我们想象的要大得多。

4.10 人工认知主体

人工认知主体，或曰人工认知系统，说得通俗一些就是具有和人同等智能的机器，更形象地说就是象一些科幻（SF）作品所描述的那种很聪明的机器人。人工认知主体是否能够实现，能不能设计和制造出和人一样聪明的机器人，是一个容易引起从哲学、科学、技术等多种角度进行争论的问题。与此类似的问题还有很多，例如“机器能不能思维？”，“人是机器吗？”等等。这类问题曾经引起过广泛的争论，但是由于争论的任何一方都很难提供充足的证据，结果陷入无结果的空谈，所以现在大多数人都对此类

问题敬而远之。

从普遍意义上讲，证明任何事情之不可能都要比证明其可能困难得多，证明可能只要作出一个实例即可，证明不可能则需要对无限时间和无限领域的推测。在某个科学发展水平上，在某个范围之内被证明不可能的事，在新的水平和新的领域里发现是可能的，这种事例在历史上是很常见的。例如在化学领域中一种元素不可能变成另一种元素，但在核物理学中是完全可能的。关于人工认知主体，迄今为止还没有见到过一个关于其不可能的有说服力的证据，而关于其可能性，本章已经提出了一个具有工程可实现性的理论框架，或者说是一个具有可证伪性的理论。这一基础理论可以作为实现人工认知主体的可行性论证，也是进一步进行更具体的研究和设计的出发点。虽然这一理论的最终评价有待于未来的研究和验证，但就理论本身而言，已经对与认知有关的各种现象具有相当强的解释能力，而且具有从哲学、生物学到认知科学的广域的内部一致性，体现了理论的深度。至少，对于传统人工智能的一切有力的批判，例如 Dreyfus 的批判，对于本书所述的理论都不会构成威胁，甚至反而是对本书观点的支持。

在可行性明确之后，研究、设计和试制人工认知主体具有理论和应用两方面的重大意义。在理论方面，只有我们能够根据某种理论制造出和人的智能不相上下的人工认知系统，我们才可以认为这种关于认知的理论是足够明确的和有说服力的。换句话说，如果我们不能制出具有感情、意识和高度智能的机器，就说明我们对认知过程的理解是不充分的，这是对理论的高标准要求。即使想要得出人工认知系统不可能实现的结论，也只能以可能实现作为研究的出发点，因为如果以“不可能”作为出发点，任何有价值的研究都无从开始，只能导致纸上谈兵。当然，我们认为人工认知系统可能实现是有充分根据的，人脑这种天然认知系统本身已经构成对人工认知系统的存在性证明。从应用的角度看，要

想制造出能够自行学习的具有高度智能的实用系统，为了回避指数爆炸，也必然需要利用广义进化范式，使系统按进化和逐层建构的方式工作，为此也必须使系统具有某种程度的感情评价和意识软硬结构的特征。而研制人工认知系统，将为设计高智能的其他应用系统提供具体的理论和技术支持。

关于制造人工认知主体，还存在着一些非技术性的问题，主要是社会心理方面的问题和伦理学方面的问题。社会心理方面的问题主要表现为一种“科学恐惧症”。对于未知的事物产生畏惧的情绪，是人和具有一定认知能力的动物共通的保护性本能。但是这种情绪所导致的健康表现应该是一种谨慎的态度，而不应该是盲目的反对和逃避。科学总是在创造前所未有的新事物，而且由于科学的逐层建构的广义进化性使得掌握科学知识往往需要长时间的学习，因而难以普及到大多数人，这就成为科学恐惧症产生的社会心理学基础。汽车、飞机和核电站都有可能发生事故，但是由于其巨大的经济效益使我们只能去尽量减少事故，而不可能禁止汽车行驶。中国古代寓言《黔之驴》就是这种认知过程的典型事例。老虎第一次听见驴叫，也会产生恐惧心理，但是老虎通过谨慎的试探而了解驴子的能力，最终可以把驴吃掉，如果老虎听见驴的吼声就逃之夭夭，那么永远也吃不到驴肉。人造的未知事物与天然的未知事物有所不同，人造的未知事物在制造之前至少其制造者对于这一事物具有相当充分的理解，对其可能产生的危害和利益也能有一定程度的估计。随着新事物的广泛应用而逐步被多数人理解，人们自然会创造出避免其危害的对策。要求汽车的发明者完全预测出现代化城市交通的事故死亡率当然是不现实的，但是汽车的设计者对于汽车的时速等基本性能应该是很清楚的。就我们目前所具备的关于人工认知主体的理论知识，也可以对于人工认知主体的性能作出一定程度的预测。可以想象，一个具有与人脑相同工作原理的人工认知主体有可能看起来比人更

聪明，因为人工认知主体可以摆脱人脑所受的生物学约束，例如神经元的处理速度及脑规模的生理学限制。假如一个人工认知系统和一个人脑在信息处理过程方面完全等效，而人工认知系统比人脑的基本运行速度高出一倍，使得人用十分钟解决的问题机器只用五分钟就能解决，那么从外观效果上我们会觉得机器比人更聪明。由于摆脱了生物学制约，人工认知主体还可能具有一些人脑所不可能具备的性能，例如经过学习达到很高智能的人工认知主体可以很方便地进行大量拷贝，复制天才是很容易的。另一方面，人工认知主体虽然能够摆脱生物学的束缚，但依然受到信息理论的制约，例如无法突破短桥原理。因此，机器虽然可能表现得比人更聪明，但也不会象神鬼妖魔那样不可理解，而仍会是通情达理的。由于人工认知主体是基于评价的，而评价准则的基础不可能离开存在、经济和预测原则，如果违反这三大原则，机器的表现将会是愚蠢的而不是聪明的。由于机器与人不具有共同的生物学基础，所以某些层次上的价值观会有些不同，但是这种差异也不可能太大，一个原因是由于二者都建立在相同的基本价值观之上，另一个原因是机器的智能在相当长的时期里要依靠人类的文化积累，而不可能独自建立起来，因此机器与人类的价值体系相差过大对于机器智能的发展是不利的。由于是在人类文化基础上进行建构，而机器也遵循短桥原理，所以机器所作的创造发明仍然是人类可以理解的。因此我们认为高度智能的机器的表现会是通情达理的，而不会是一群不可理喻的妖魔鬼怪。

科学与伦理学的关系一向是一个难以取得一致意见的领域。医学和基因工程学的每一次进步几乎都会受到伦理学的纠缠，而科学中的伦理学问题往往导致无结果的争论，最终似乎只能依赖科学家的良心。对于制造有感情有意识的机器来说，伦理学的争论恐怕也是难以避免的。科学与伦理学之间产生冲突的根本原因在于二者的基本出发点是不同的，科学崇尚理性而反对形而上学，

但迄今为止的伦理学主张往往立足于形而上学的基础之上，具有浓厚的宗教背景。不具备共同基础的争论通常是无效的，就象争论欧氏几何与非欧几何的定理谁对谁错一样的无意义。为了使科学与伦理学之间进行有效的讨论，需要使二者立足于相同的基础。由于科学与形而上学之间的根本性对立，科学的讨论不可能以形而上学为基础，否则科学本身将不复存在，因此解决这个问题的唯一途径就是为伦理学建立非形而上学的基础。由于在本书中已经建立了非形而上学的价值观体系，而伦理学的集中表现也就是一组价值观，因此建立非形而上学的伦理学基础并不是不可能的。本书的第五章将要涉及这方面的问题。从比较文化学的角度看，基督教文化的深层精神结构由于其绝对性和形而上学性，比较容易产生伦理学的争论，哪怕象避孕这样简单的问题也会引起形而上学的争端和教义的冲突，但是在儒家文化背景下很多与科学有关的伦理学问题根本不会产生，即使产生了伦理学问题，在价值相对比较的基础上也比在形而上学的基础上容易解决得多。

第五章 广义进化系统

5.1 复杂巨系统

科学理论的统一性是科学活动所追求的重要目标之一，高度统一的理论也意味着深度和稳定性。牛顿在追求苹果和月亮的统一之中创立了力学的深层原理，爱因斯坦为克服牛顿力学与电磁学之间的不统一而创立了相对论，并且为追求更深层的统一场论而付出了后半生。自然科学与社会科学之间的分裂是现今依然存在的现实，弥合这种分裂仍然是当今学者所感兴趣的课题。我国学者就曾经倡导用系统工程的方法研究社会科学，并提出了复杂巨系统这一重要概念。由于社会作为一个系统是极为复杂的，通常的系统论在处理这样复杂的系统时往往感到力量不足，只能处理一些局部的、小规模、或是经过极度简化的模型。这也就是复杂巨系统理论之所以重要的原因。为了建立复杂巨系统理论，我们首先需要明确复杂巨系统与传统的系统论所研究的系统之间有什么不同，然后才能考虑所用的方法有什么不同。

顾名思义，复杂巨系统的两个明显特征是复杂和巨大。如果把系统定义成相互作用着的诸元素的综合整体，那么元素的个数可以作为系统巨大程度的度量。但是元素数量的巨大并不是构成复杂巨系统的充分条件，巨大的系统未必都是复杂的。有一种关于复杂性的定义，它把复杂度看成是系统中相互关系的数量与元素数量减一之比，这种复杂度定义是严格量化的，但与我们关于复杂系统的概念不大一致。例如一杯水中的分子数量是巨大的，一个分子与很多其他分子有力学关系，但是我们并不把一杯水看成

是一个复杂巨系统，其理由是所有分子之间的关系都可以用相同的原理来进行解释和描述。从这个例子可以看出，我们关于复杂性的理解还依存于系统原理和机制的复杂性。这种原理和机制的复杂性可以分为纵向和横向两种，纵向的复杂性表现为对于某些系统我们无法象对待一杯水那样用少数的原理和机制去解释大量元素之间的关系，而需要用逐层分析的方法把大系统分成子系统，每个子系统又分成更低层次的子系统，才能弄清其原理和机制，而各层次之间的原理又有所不同。横向的复杂性则表现为在相同层次的不同子系统之间也存在原理和机制的差异。例如在以细胞为元素的人体系统之中，要从纵向分为系统、器官、组织、细胞几个层次，而从横向上看，各个系统，例如消化系统和神经系统之间，各个器官，例如胃和肠之间，都存在着不同的原理和机制。作为相反的例子，电话网则不具备这种纵向和横向的原理复杂性，虽然世界规模的电话系统可以包含几亿个用户，元素数量是巨大的，而且也具有国际、国家、地区、局、用户这样的多层结构，但是除了局对用户这一层与其他上层在原理上略有不同之外，各层次在纵向和横向都保持大致相同的原理和机制，因此电话网看起来是个巨大但却简单的系统。

以上我们从元素间相互关系的数量和原理的数量这两方面探讨了复杂性的度量，但是仅从这两方面还不足以区分复杂巨系统与通常系统在概念上的不同。例如一个运行航天飞机的系统，无论从零件数量和零件之间相互关系，或者从所涉及的原理和机制的数量上看，都是复杂而巨大的系统，但是与社会这样的复杂巨系统相比，其复杂程度还是有明显的不同(或区别)，我们可以很容易的判断一架航天飞机是否偏离了正常的轨道，但是很难判断一个社会是否正在正常地发展。二者的一个重要区别是，对于不论多么复杂的工程系统，其评价准则都是从外部给定的，而哪怕是一个小规模的社会，其评价准则都来源于系统内部。

用广义进化论的观点来看，我们可以对复杂巨系统给出一个并非仅仅从数量上与普通系统相区别的定義，即把包含广义进化机制的系统称为复杂巨系统。或者说，经典的系统论是建立在欧几里德范式之上的，而复杂巨系统理论的基础是广义进化范式。从这个定义出发，我们就可以发现复杂巨系统与经典系统之间的共性和区别，看到复杂巨系统的特征和有别于经典系统论的研究重点。复杂巨系统除了具备经典大系统的元素数量巨大、元素之间关系数量大、涉及原理多等大系统的共性之外，还具有经典系统论所未能包括的特性，即源于自我表述的复杂性。对于以生物进化，文化进化，人脑学习，社会发展为典型例子的复杂巨系统，由于其具有基于内因的自进化能力，包含明显的创造性，需要用广义进化论的观点，划分软硬结构来进行研究。对于硬结构的部分，通常可以采用经典系统论的方法研究其稳定性及效率、功能等经典问题，但是广义进化系统的硬结构并不完全等同于经典系统，因为还要包括与软结构的相互作用，也就是对软结构的支持和来自软结构的建构。

对于软结构的研究来说，其研究方针与经典系统论有很大的不同，由于软结构和软硬结构的关系之中包含创造过程，从而具有原理上的不可预测性，因此对于软结构来说稳定性、效率、功能分析这样一些经典系统论中经常遇到的问题变得不那么重要，而研究的重点则变为由进化机制而产生的问题，这些问题在经典系统论中是不会遇到的。正是广义进化机制使得复杂巨系统具备创造性，成为开放的复杂巨系统。进化的基本机制是变异与选择，选择意味着价值体系的重要性，而变异则意味着多样化的价值。变异所要求的多样化的价值在经典系统论中是不存在的，选用最佳的零件按照最佳的设计经过最佳的工艺过程制造的电视机是最佳的产品。但是如果能够选择最佳的遗传基因复制出大量完全相同的最佳的人种，再让他们接受统一的最佳教育，由这种最佳克隆

组成的社会却肯定是一个不佳的社会。因为失去了变异也就失去了进化的可能性，这种克隆社会一定是最容易灭亡的社会，正如由完全相同的基因组成的生物种群一样，不可能对进化压力作出任何反应，肯定是最容易灭绝的。这种观点从一个方面揭示了“理想主义的谬误”。从建立在欧几里德范式的牛顿力学来看，一个理想的机械传动装置应该是无摩擦的，具有百分之百的效率而无磨损，我们可以采取各种巧妙的技术去尽可能减少摩擦而逼近理想状态，制造出更好的机器。但是如果我们用相同的方法使一个社会趋向理想化时，可能恰恰造成了社会或社会形态的灭亡。对于“通往地狱的道路上铺满了善良的愿望”这句名言，也可以作出其逆定理，即“通往天堂的道路上充斥着邪恶的愿望”。当然这不意味着我们的行为应该为恶而不行善，只是说明对于善恶的判断并不是一件很简单的事情，有时候认为是至善的行为却会导致恶的结局。

从选择的角度看，价值体系的研究是软结构的核心问题，任何一种选择都不可能不依据某种价值判断。由于软结构的丰富的创造性，局部的和具体的功能、行为、结构等在通常系统中重要且稳定的因素在软结构之中都属于被创造的对象，成为从原理上不可预测的，同时也是不稳定的因素。因此对于软结构来说，价值体系才是处于更深层次的起主导作用的相对稳定的因素，从而也是更重要而比较值得研究的更深入体现软结构特征的因素。具体的理论、技术、方法、习惯、生活样式等都属于文化的产物，只有价值体系才是文化的核心。文化的表层产物是通过价值体系过滤的结果，文化产品在不同文化体系之间交流的成败取决于价值体系的选择性过滤，异种文化之间的冲突从根本上来说是价值观的冲突。当然，价值体系的稳定性也不是绝对的，特别是表层的价值观依然处于进化过程之中，但是只要评价准则可以用于文化产品的选择，其稳定性就必然高于具体的文化产品。表层价

价值观的进化仍然是在深层价值观的选择下实现的，因此越是深层的价值观其稳定性就越高。

本书的二、三、四章从哲学的高度建立有关价值体系的基础理论，并在认识论、进化生物学、认知科学等领域应用和丰富了价值理论，本章则通过社会科学的一些实例来进一步展开价值分析的方法和原则。

5.2 全局观点与分离主义

评价只在相对比较中才有意义，评价结果与目标及环境有很大的关系。环境因素之中很重要的一个因素是评价的时空领域，即在时间上近期与长远的利益关系和空间中的局部与全局的利益关系。多人囚犯困境的哈丁解是关于局部与全局利益的冲突问题，而阿克塞洛德解则是处理近期与长远利益关系的。系统中各个子系统可能存在着相互矛盾的价值取向，对于这种问题有时不得不采取一种全局最优的（或是各方都可接受的）妥协方案而在某种程度上牺牲局部利益，但是一个设计得更合理的系统则可以巧妙地分离子系统，使得各子系统之间对立的價值取向可以各行其事而互不干扰，在各自局部最佳的同时求得全局最佳。广义进化系统中软硬结构的分离就是这种分离的一个例子，使得系统效率与创造性之间的矛盾要求在分离之中各得其所。这种由分离产生的利益往往发生在大发明之后，例如从腔肠动物向“通肠动物”的进化过程中，肛门的发明导致了消化道的前后之分，为胃和肠的分离创造了条件。然而在文化进化的过程中大发明的意义往往不能充分为人们所认识，由于传统的力量而使已经不合理的结构得到维持，这就使对于分离主义的强调有其现实意义。当然，分离主义并不是一个新奇的想法，大系统中的解耦问题一向是受到重视的，而解耦就是一种分离。

由于分离不充分而使系统未能达到最佳状态的一个典型例子，是汉字文章的排列问题。众所周知，我国现行的从左到右横排的书写和印刷规则是50年代文字改革的一个成果，在此之前，汉字文章排列一直是竖排从右到左的老习惯，时至今日，我国港台地区和日本等其他使用汉字的国家仍然沿用竖排和从右到左的方式。与竖排从右到左相比，横排从左到右有着显而易见的优越性。从阅读的角度看，人的双眼是水平排列的，眼睛水平方向的转动角度比垂直方向大得多，因而阅读横排文章只靠眼动扫描即可，而阅读竖排文章则要加上头部的俯仰运动，引起颈部疲劳。从书写的角度看，由于多数人用右手书写，从右到左的排列一方面与汉字从左到右的笔顺相逆，在横写时造成多余的回笔，而在竖写时手的位置正好挡住前文，影响文思，甚至袖子会擦坏未干的墨迹。但是我们的祖先为什么会选择这种不合理的排列方案呢？我们的祖先并不比我们笨，在他们选定竖排从右到左的方案时是有其充分理由的。在纸张发明之前，汉字主要是写在竹简或木简上，竹简是长条形的竹片，每一片写一两行字，然后用绳子从简的两端把很多简栓成一串构成一册书，“册”这个汉字就是把竹简穿在一起的象形文字，一长串的简在收藏时要卷起来，所以“册”和“卷”至今还是我们用于书籍的量词。在阅读竹简穿成的书时，如果字是横向排列的，那么成卷的书要放在肚子附近，读过的部分则要放在前方一定距离，这样操作是非常不方便的。因此竹简只能是竖方向放置，使阅读时放卷和收卷在左右两侧，操作比较容易。由于大多数人是右利手的，而收卷比放卷更费力，因此左手放卷右手收卷是比较合理的，所以行的排列也自然就应该从右到左的了。由于用竹简写书是先写后穿成册的，编辑与编织使用相同的“编”字就反映了这个过程，所以在写竹简时竹简是单独一片片存在的，写好后可以放在任何位置，不会发生遮挡或袖子擦坏之类的问题。所以在竹简时代竖排和从右到左是充分优化的

结果，具有充足的合理性。竖排和从右到左的不合理性出现在纸张发明之后。

纸张的发明是一个大发明，把竹简这样的一维信息媒体推广成平面的二维媒体，扩展了一个自由度，使书写排列变成可横可竖的了，如果说对于沿用竹简的卷轴方式来说竖排从右到左还有一定的合理性，那么在成页装订的书出现以后，竖排从右到左已经是完全不合理的了。我们祖先的失误就在于，当一个大的发明出现之后，未能充分利用大发明给我们带来的自由度，一味维护传统，迁就习惯，导致两千年非最佳状态的沉重损失。时至今日，重温这段历史，依然是非常重要的。特别是计算机的发明，其意义远在纸张之上，充分认识新媒体带来的自由度，充分发挥这种自由度的优势，可以避免重演两千年前的失误。有趣的是，计算机存储器的编号又回到了一维的，而屏幕显示又回到了卷轴（滚屏）的形式，但是一维的存储器可以提供任意多维的存储空间，而卷轴也由多种检索方式提供了比分页更好的任意检索性能，克服了过去卷轴随机检索不便的缺点。

计算机给我们带来的自由度当然远不只是存储和显示，对于中国人来说，计算机技术的发展可以说是开辟了汉字的新纪元。汉字是世界上仅存的一种仍被广泛使用的象形文字。与拼音文字相比，象形文字的优点是便于阅读，缺点则是不便于书写。象形文字在阅读方面优于拼音文字这一点，可能被熟悉拼音文字的人认为是一种偏见，因为阅读的效率显然与熟悉程度有密切关系。对于拼音文字与象形文字在阅读效率方面的对比，笔者曾经作过一个简单的实验，用日文作为测试文字，因为在日文中同时使用汉字和假名两种文字，汉字是象形文字而假名是表音文字。实验的方法是在列车高速驶过车站时，让被试者读取一闪而过的车站站牌上的站名，其结果是无论被试者是中国人还是日本人，在读站牌上的汉字时都能正确读出站名，但读假名时往往不成功。站牌

上汉字与假名字体一样大，同一个车站的站名代表着相同的信息，日本通常的教育中学习假名比汉字要早几年，日常阅读的文章中假名多于汉字，而假名的总个数是常用汉字的二十分之一，因此日本人接触同一个假名的频度远远高于同一个汉字的频度，对假名的熟悉程度高于汉字。尽管如此，在读站牌的实验中仍然是汉字的读取效率高于假名，就可以充分证明象形文字有利于阅读这一结论。至于象形文字不利于书写，我想是每一个学过汉字的人都会深有体会的，上小学时我们不害怕被老师叫起来读课文，但一到听写的时候却不免心惊胆颤，一个字一旦认识就永远认识，但以往会写的字却会突然想不起来怎么写了。这样的问题在使用拼音文字时不会发生，象英语那种拼写与读音不一一对应的语种还可能发生拼写错误，而德语这种读写一义性很强的语种除了笔误不太可能出错。正是这种优劣互见的局面，使得汉字这种象形文字不被淘汰，初期学习时多付出的成本和书写的困难，可以在阅读效率中得到补偿。

正如通肠动物的发明为胃肠分离提供了可能性，纸张的发明为字的排列方向与媒体方向的分离提供了可能性一样，计算机的发明为阅读与书写的分离提供了可能性。在计算机发明之前的纸张文化之内，书写的字体就是我们阅读的字体，二者是完全一致不可分离的，而利用计算机书写，写和读的“字体”可以完全分离，例如可以用键盘上击键的方法写出象形文字来，而文字的质量与击键的熟练程度无关。读写分离可以为文化体系带来莫大的利益，这是由于在人类的模式生成与模式识别之间存在着相互对立的关系，我们称之为模式生成与识别的“互补定律”。“互补第一定律”是指模式生成与识别所费的代价总量不变，或者说当模式生成花费的代价较大时，识别所花的代价就减少，反之生成的代价减少时，识别的代价就会相应增加，例如工整的字体写起来费力费时，读起来就比较容易，而潦草的字迹写起来省事，读起

来就费劲多了；电台播音员的话很容易听懂，但说起来很累，公共汽车售票员按最省力的方式说话，听起来最困难。“互补第一定律”可以用进化的机制加以解释，即人的相互通讯总是在进化过程中通过变异和选择找到最经济的形式，达到模式生成和识别的代价总和极小，生成与识别代价都很高的表达方式会被淘汰，最后剩下的就是生成与识别代价的折衷产物，一方增加则另一方减少。“互补第二定律”讲的是人机互补关系。人和计算机相比，人的模式识别能力比计算机强得多，而计算机在模式生成方面则比人要强得多。对人来说，学会识字比学会写字要省力得多，而对计算机来说则恰恰相反，识字比写字要困难得多。其原因就在于，人脑中用于模式识别的神经资源远远多于用于模式生成(运动)的部分，而计算机在模式识别方面的方法和资源上都远远没有发展到能与人脑匹敌的水平。反之在模式生成方面，由于人具有学习能力，所以必须不断产生变异以满足学习的要求，这就使得人在模式生成的过程中会有较多的“差错”和“变形”，而且在学习过程中由于经济原则的作用，使得模式生成自发地向省力方向发展，所以人在无意识之中字迹会日趋潦草，说话会产生含混和连音省略等现象，而目前的计算机由于不具备学习能力，模式生成不存在退化机制，因此计算机“写”出来的字总是工整划一，不会偷懒的。

由于“互补定律”的存在，计算机所实现的“读写分离”可以使文化体系中基于文字的信息传递方式在新的水平上进一步实现优化，综合拼音与象形文字二者的优点，结合人与计算机双方的优势，创造出更高水平的汉字文化。这种无纸文化的构成是，输入靠拼音，显示用象形，传递和存储则用计算机编码。在这种文字体系之中，既保留了汉字作为象形文字的便于阅读的优越性，又加进了拼音文字易学易写的优点。小学生只要求认识汉字即可，并不需要会写汉字，这样小学的语文教育至少可以节省一半的时间

和精力。这样一个提案很可能遭到传统主义者的反对：“我们怎么能培养出一代不会写字的文化人呢？”这里需要作一点说明：随着社会上的技术进步，总是不断会有一些过去被认为是基本生活技能的技术被淘汰。对于原始人来说，钻木取火是必须掌握的生活技能，而现代人很少有人会这一手；用惯了气体燃料的孩子们，不象上一代人那样会生煤炉子；在计算机时代，写字和珠算都是将要被淘汰的技术。文化是进化的系统，应该重视多样化的价值，由少数爱好者去钻研书法和珠算作为保存文化基因来说是非常有意义的，只是不必让所有人浪费时间去学习即将被淘汰的技能。作为少数人的爱好，研究钻木取火和制作打制石器都是很有意义的事，可以让我们懂得我们的祖先有多么了不起。

关于汉字输入，我国已有数百种方案，但从原理上分，不外乎利用字形、读音和字义作为输入的关键信息。字义由于定义的随意性大，记忆困难，因而采用者不多，即使采用也只是放在辅助地位，关键信息仍是字形和读音两种。表面上看，字形输入方案和读音输入方案各有优缺点，一时似乎难分胜负。字形输入的优点是不受方言限制，缺点则是学习和记忆都比读音输入困难，读音输入的优点是易于学习和记忆，缺点则是不会说普通话的人不易掌握。但从长远和全局的观点看，读音输入与字形输入相比具有绝对的优势，其原因是读音输入不需要使用者会写字，只要能识字即可，从而使我们获得简化文字教育的巨大利益，而字形输入则没有这个优点，因此将来必定会被淘汰。读音输入的方言问题是可以解决的，我们可以提出两个方法，一个是推广普通话，把计算机的普及作为推广普通话的促进因素；另一个方法是为每一种方言单独设计一个与该方言适合的读音输入方案。采用哪种方法在某种程度上取决于国家政策，但从可行性上看，后一种方法显然比较省事，遗憾的是目前关于方言发音体系的研究还很不够，没有达到象普通话发音体系那样具备完整的符号系统。

读写分离给汉字文化的发展带来了前所未有的自由度。现今汉字的字形结构是由古代象形文字演化而来的，演化过程主要是受到认读方便、书写方便和书写工具多方面的影响，是一些相互矛盾要求的折衷产物，因此从任一方面看都不可能充分优化。例如由曲线组成的象形文字比用直线组成的楷书要易识得多，但书写却比较困难，因而逐步被楷书所代替。在计算机所造成的读写分离的环境中，汉字作为只读不写的文字，完全可以不考虑书写的难易，只考虑认读的方便，可以向更象形的方向发展。其结构可以不限于直线，可以采用曲线、涂黑领域、甚至浓淡和彩色来表示，字的复杂程度不受书写限制，可以就是一张小图，如同计算机上显示的图标，这样的象形文字既便于识别也便于学习，看图识字变成了看图识图。当然，为了设计一套高度合理的象形文字，还需要进行多方面的研究，例如视觉心理学的研究。另一方面，读写分离使我们可以设计一种专门用于书写的拼音文字，作为向计算机输入用的文字。目前的计算机普遍使用键盘和鼠标作为输入设备，不需要为输入设计文字。但是随着计算机体积的逐步小型化，键盘将不再适合用作输入，只能使用笔触式的点坐标输入的方法代替键盘和鼠标。此时如果用在线识别的方式直接输入汉字，不仅可靠性差而且输入效率很低，也不能发挥计算机的优越性实现不用学写字的汉字文化。因此需要设计一种专门用于书写输入的拼音文字，力求简单易学，易写高效，以及在线识别的高可靠性，但对于阅读方便则可以毫不在意，类似于速记文字那样，因为这种文字只用于写不用于读，进入计算机后就被变换成象形文字的编码，用象形文字显示。

关于分离主义的另一个实例是在数字时代的著作权问题的根本解决方案。现存的著作权保护的基本机制几乎都是在某种程度上依靠复制的困难。例如用复印机复印书籍的成本通常要高于印刷成本，录音带和录像带在复制之后会使音质和像质变坏，这些

都是对著作权保护有利的因素。但是随着计算机技术的发展，复制困难的堤坝正在日益走向崩溃，计算机软件由于其容易复制且价格高昂而率先成为著作权受侵害最为严重的信息产品。计算机带动了数字技术使得文字、声音、图形、图像、动图像等各类信息的生产、传送、存储、处理、使用都向数字化的方向发展，多媒体技术就是这一发展方向的综合表现。随着数字化的推进，信息产品在生产、流通、存储、处理等各方面都大大提高了效率和质量，同时也导致复制困难的著作权堤坝进一步崩溃。普通的录音带或录像带每复制一次都会造成失真和噪音的增加，而数字的录音带却可以多次复制而不会造成信息损失，这样就对以复制困难为基础的著作权系统造成很大的威胁。在原有的著作权系统之中，为了保护著作权不得不采取人为的手段增加复制的困难，例如对计算机软件进行加密，或者象有线电视和商用数据库那样采用个别契约的封闭方式提供信息，甚至对硬件进行限制。在家用数字录音机问世时，由于著作权组织的压力，不得不在设计上加入限制复制的措施。所有这些措施，结果都是降低信息流通效率，增加信息流通成本，不利于系统全局优化的。而著作权保护与信息流通效率之间的矛盾并不是必然产生的，仅仅是著作权保护与复制难度之间的传统关系所造成的。如果采用分离主义的态度，设计新的著作权保护系统，把著作权保护与复制难度彻底分开，就可以避免二者之间的矛盾，设计出著作权保护更为可靠合理，而信息流通成本又能达到极小化的综合信息系统。

数字信息的特点之一是易于复制，这一特点是与以复制困难为基础的著作权保护体制相矛盾的。但另一方面，数字信息还有另一个重要的特点，那就是数字信息的间接性。写在纸上的字可以直接看到，而存在磁盘中的文章则必须经过计算机的加工才能在屏幕上显示；普通的录音带可以直接放出声音，而数字录音则必须经过一定的解码（数模转换）过程才能变为通常的音频信号。

如果利用这种间接性来设计新的著作权保护体系，可以使信息产品的流通体制与著作权保护完全分离，信息与信息载体完全分离，从而使得信息产品的流通体制可以不受著作权问题的干扰而充分优化，最大限度地降低信息流通成本，同时又能使著作权得到更完善、更合理、更精确的保护。

作为这种新型著作权保护体系的一个实例，我们提出一个“多信息精密结算系统”（*Multiple Information Accurate Account System*, MIAAS）的构思以利于具体分析，其中心思想是用一个共同的系统保护所有信息产品的著作权，使得整个系统运行成本很低，可以产生巨大的经济效益。所谓信息产品，包括计算机软件（开发用软件、各类应用软件、游戏软件等）、各种出版物（报刊杂志、书籍、画册、地图等，其中含有文字、图形、静止图象）、音像制品（音乐、歌曲、影视节目等）、以及综合以上各类信息的多媒体信息产品（例如含有主控制程序、文字、图片及音像制品等多种信息形式的教育软件）。“多信息精密结算系统”主要由以下四个部分构成：信息生产者接口、信息消费者接口、开放式信息流通系统、结算系统。

信息生产者接口是MIAAS与信息生产者之间发生关系的部门，其主要作用是接收信息生产者所生产的信息产品。信息生产者所生产的原始信息，例如记者写的报道及拍的照片、评论员写的文章、作家写的小说、歌手演唱的歌曲、制片人拍摄的影视节目等，都以数字形式（影视节目也可以用模拟形式）的媒体（例如磁盘、光盘、磁带）方式送到生产者接口部门，象文字信息这种信息量较小的产品，也可以通过电话网传递。原始信息在生产者接口部门要经过以下的加工处理：对每一件信息产品规定一个统一的编号，决定单位价格，决定流通方式及流通成本，决定检索信息（分类和关键词等），决定信息性质及消费者资格，登记著作权所有者姓名及银行帐号，然后将上述各参数传送到结算系统

储存备用；对原始信息产品附加上编号、单价等信息，以一定的方式编成密码形式，成为流通形式的信息产品，送交开放式流通系统，并从著作权所有者帐号上收取基本流通费用。由于信息产品已经密码化了，流通方式就与著作权完全分离了，因此流通可以是完全开放的，可以通过广播的方式，也可以是个人之间复制的方式，都是完全自由的，从而可以最大限度地优化，达到最低的成本，这就是开放式信息流通系统的特点。消费者接口的核心部分是“计量解码器”，其功能是在将密码信息翻译成信息消费者可用的信息形式的同时将消费者使用的信息产品和用量记录下来，即记录下信息产品的编号和用量。用量的计量单位随信息种类而有所不同，静止信息（例如文字、图形、照片）可按字数或篇幅计量，音像制品可按时间计算，功能信息（例如应用软件或游戏软件）可按使用次数或时间计量。“计量解码器”定期与消费者接口的终端进行联系，将该消费者使用过的所有信息产品的编号和各产品的用量传送给消费者接口进行结算，同时从消费者接口终端获得下一时期的解码用密码键。“计量解码器”的具体结构可以有两种形式，一种是在各种信息终端（电视、音响、计算机）上都装有计量解码器接口，计量解码器则作成集成电路卡的形式，使用时插在相应的信息终端上，结帐时插到消费者接口终端上进行结帐。另一种形式则是信息消费者持有一台内部装有计量解码器的多媒体信息机，该信息机包含了计算机、音响、电视机、电子邮件、传真机、电话机、电子手册等全部功能，可以自动按期通过电话网与消费者接口终端联系，完成结帐手续。消费者接口终端得到消费者结帐信息之后，按各信息产品编号分别累计信息用量，定期将累计结果传送给结算系统，结算系统进一步按各信息产品编号累计信息用量，定期按照各产品累计用量和单价计算出该产品该期间的收益，扣除系统成本和一定的系统和流通利益之后将收益转到著作权所有者的银行帐号上。

多信息精确结算系统(MIAAS)与传统的信息发行系统相比,具有高效率(高经济性)、高精度(高合理性)、高柔性(高通用性)等明显的优越性。高效率主要体现在高速度和低成本两个方面。由于著作权保护与发行方式分离,可以采用最高效的信息传播方式而不必担心著作权得不到保护。以报刊杂志小说等文字信息为例,由于其信息量小,通过广播的方式传播信息可以获得极高的效率。例如利用电视台夜间停播时间传送报纸,由于电视频道具有很宽的频带,可以用于高速传输数字信息,按保守的估计可达到1M bit/秒,相当于每秒至少可以传递5万汉字的速度,传送一份报纸只需数秒即可,传送长篇小说也用不了一分钟时间。用这种方式在全市发行一份报纸的成本只相当于电视台发射部分运行几秒钟的费用。假定发射台功耗为100kW,发射一份报纸用3.6秒,则耗电量只有一度,电费只有几角钱,加上设备折旧和维护等各方面费用,总成本不会超过百元,但是用传统方法发行报纸,排版、印刷、纸张、运输、分送等各个环节,哪一项的成本也远远超过百元。从速度上看,广播报纸时间只用几秒,而通常的报纸发行全过程不会少于几个小时,广播一部长篇小说只要十几秒至几十秒,而通常的出版发行一般需要一个月以上。对于影视作品来说,由于信息量大,利用广播的方式成本降低不很明显,但是在考虑著作权保护的情况下,与有线电视(Cable TV)系统相比较,在系统成本和著作权保护的精度以及信息消费者的灵活性等方面,都更为优越。

MIAAS在著作权保护的高精度方面也远远优于传统的体制,无论对于信息生产者和信息消费者来说,对于著作权的保护都更为充分、更为合理。例如我们每天看报纸并不是所有文章都看,而是选择我们感兴趣的一部分消息和文章,但是对于记者及其他在报上写文章的人来说,读者的这种选择却不能直接反映出来,对于每篇文章的评价只是编者的直觉估计,而在MIAAS系统之中,

每篇文章可以分别计算读者数量和收益,在收益分配上可以避免大锅饭的不合理性。因此 MIAAS 也解决了信息的原始生产者和信息的二次生产(组织)者之间的利益分配问题。例如在一本小说集里有不同作者的作品时应该如何合理分配收益的问题,在传统的著作权体制中是很难解决的。这个问题在电子出版系统中越来越尖锐,如果我们把五部水平相差不大的中篇小说编在一部小说集中,收益分配问题还不太突出,但是一张只读光盘(CD-ROM)的容量可达 500MB,相当于两亿五千万汉字,可以存放 50 万字的长篇小说 500 部,这样我们可以用一张只读光盘出版近年所有小说,此时如何合理地分配这部小说集的收益就不是个简单问题了,只有用 MIAAS 才能圆满解决。随着多媒体技术和大容量存储技术的发展,从技术上看已经有可能出版非常大规模高功能的多媒体信息产品,但是由于这种大规模信息产品只由少数著作权持有者生产是不可能的,需要编辑现存的著作权极为分散的产品,因此使得与所有著作权持有者分别签订合同成为现实上难以完成的工作量,一些高水平的信息产品也因此难以问世,而 MIAAS 可以对一个媒体中不同部分信息分别对著作权费用进行核算,因此可以完全消除这种障碍。从信息消费者的角度看,MIAAS 也可以使其对著作权的支付更为精确合理,如果一部小说只看了一章就不想再看下去,那么只需要支付一章的费用,一部电影只看了 10 分钟就只支付 10 分钟的费用,一个计算机软件只使用一次就只支付一次的费用。这种精确性对于解决计算机软件的著作权问题来说极为重要。以往的计算机软件发行体制对用户来说是一次付清全部著作权费用,因而显得十分昂贵,用户一旦选购了某一种软件,则即使再出现性能更好的软件也很难再投资更换,因此使得用户难以有充分理由下决心购买软件,只好采取无偿使用复制品的方法侵害著作权,而此种侵害行为又会引起软件价格的进一步提高,形成侵害著作权的恶性循环。在 MIAAS 系统之中,用户使用一次软

件所需费用很少,可以放心大胆地同时使用多种软件进行比较和选择,而在出现更好的软件时又可以放弃使用原来的软件而不受损失。这样对于软件生产者来说,著作权得到完善的保护,而对于用户来说软件使用费用又很合理,而且具有随时选择产品的自由。

MIAAS 的第三个重要的优点是其高度的柔软性和通用性,可以适应各种不同的要求。只要改变信息生产者接口管理程序,同样的 MIAAS 系统可以适应从绝对出版自由的体制到绝对中央集权体制的各种信息体制。通过资格代码控制,同一个系统既可以成为最广泛的大众传播媒介,又可以同时传送小范围的内部资料,例如实现对未成年者的信息限制,或是某个组织内部独占的分级信息传达。一个系统可以同时实现高度公开和高度机密的信息传递。MIAAS 的价格体系也是非常灵活多样的,可以按信息的使用时间计价,也可以按篇幅、使用次数计价,有些信息的价格可以随着时间的推移而减少,也有些信息可以随使用次数而逐步降低价格,按照不同信息的性质,可以设定与之相适应的计价方法。可以设定零价格,甚至也可以设定负价格,例如广告之类的信息。由于发行系统的开放性,可以选择最高速的传播手段,因此 MIAAS 系统可用于发行实时性的信息,例如股票、期货、汇率等很快变动的信息,以及交通情况和市场信息等。MIAAS 为信息消费者提供了最大的选择自由,例如我们可以不必考虑订哪一份报纸,而可以每天看到所有报纸,从中选择自己感兴趣的部分,只支付阅读部分的费用,而且可以按自己的兴趣让计算机按信息性质或关键词、作者姓名等各种索引自动选择信息,相当于按自己的爱好编一份适合自己的报纸。由于著作权得到保障,优秀的影视作品可以通过电视广播这种高效传播方式来发行,电视台被劣质节目充斥的现象也会消失,信息消费者能够以合理的代价获得所需要的信息产品。由于这样一些利益,信息消费者支付计量解码器和信息终端的代价是非常合算的。因此 MIAAS 也为计算机工业开拓了人手一台的巨

大市场。

5.3 伦理学的非形而上学基础

伦理学是关于道德或善恶的学问,伦理学所研究的是道德或善恶判断所立足的理论基础问题。道德是人们应该作什么和不应该作什么的指南,伦理学则是研究我们为什么应该接受或拒绝某种道德原则的理由。简而言之,伦理学也是关于价值观的理论。由于本书在前几章中从哲学、进化生物学、认知科学等层次探讨了关于价值的基础理论,本节主要从社会科学的角度来探讨关于价值的理论,建立与其他层次相一致的伦理学理论基础。

比较原始的伦理学是建立在形而上学的基础之上的,道德的出发点往往是某种宗教教义,或是圣人、伟人的言论。但是形而上学的伦理学基础含有很多矛盾,作为基础理论的稳定性是相当差的。首先,宗教并不是唯一的,世界上存在各种各样的宗教和教派;圣人、贤人、伟人也不只一个,古往今来,世界各地的圣人、伟人也是层出不穷;他们的言论也不能保证相互一致,宗教的教义也是各有千秋,因此以宗教教义或圣人言论为基础的伦理学是难以具有统一性和稳定性的,在不同的道德观之间的争论中,这种理论基础无法作为评判的基点。另一方面,如果站在无神论的立场上看,任何宗教教义都只是一种神话,不具备任何理论意义,而站在科学的立场上看,任何圣人、伟人的言论也都不具备当然真理的地位。

在科学史上占据了长期统治地位的逻辑实证主义对近代伦理学也产生了重大的影响。从逻辑实证主义出发,对于伦理学的基础有两种观点。一种是“非认知主义”观点,认为关于道德的陈述不是关于客观事实的描述,因此不可能根据事实来判明其真伪。另一种是“认知主义”观点,认为关于道德的陈述虽然与自然界的 facts 不同,但仍然是关于客观事实的陈述,有可能辩明其真伪。非认知主

义的看法是一种虚无主义的态度,按照这种看法,所谓伦理道德仅是宗教迷信的糟粕,因此伦理学本身不可以作为认知的对象,或者说伦理学不可能作为一种学问而存在。由于这种自我否定,非认知主义不可能成为伦理学的基础。但是非认知主义之所以能够存在,是由于认知主义本身也是缺乏说服力的,道德现象的客观性确实不象自然现象的客观性那样明显。产生这种问题的根源是逻辑实证主义的立场本身并不很可靠,伦理学所应该讨论的问题是善恶而不是真伪,以真伪作为核心问题的理论体系如果说在自然科学的某些局部还能应付的话,那么在认知科学和社会科学、特别是伦理学的领域是完全无能为力的,伦理学所需要的是关于善恶的理论基础。

认知主义与非认知主义虽然都不适合作为伦理学的基础,但作为出发点也并不是一步也走不出去,还能够有一些小的进展。在非认知主义阵营之中分为两派,一派是情动主义,认为道德陈述是在体验某种与道德有关的事实时对感情所作出的表达;另一派是指令主义,认为道德陈述是当认识到某种事实时应该如何行动的指令。在认知主义阵营之中也有两派,一派是直觉主义,认为知觉是对自然现象的描述,而与此相类比,“直觉”可以对非自然的道德事实进行表达;另一派是自然主义,认为道德陈述所表达的事实可以还原成自然的事实。显然,如果不借助认知科学的力量,这四种观点都只能停留在肤浅的阶段。如果我们不知道感情在认知过程中的地位及其评价作用,不知道指令的感情评价基础,不知道直觉从何而来,以上建立在感情、指令、直觉基础上的伦理学都无法取得进展,而且不通过感情评价这一环节,道德陈述也很难被还原成自然的事实。而一旦认知科学可以对感情的理由、指令的依据和直觉的来源作出解释,上述四种观点实际上是殊途同归的。

关于伦理学的基础还存在着其他观察角度所看到的分歧。例如考虑善恶评价最基本的单位,即善恶首先是对谁而言时,可以有

两种不同的看法,一种是个人主义的,认为利害计算的基本单位应该是单独的个人;另一种是全体主义的,认为利害的计算应该是对社会全体而言的。此外,如果把道德看成是文化进化或社会进化的产物,同时对于个人来说是教育与学习的结果,那么对于这个过程也有两种看法,一种是相对主义的看法,认为社会的道德观念或个人的道德修养取决于该社会发展的历史或某个人成长的环境,因此不同的社会及文化之间或是不同个人之间必然存在道德观念的差异,而不存在普遍的道德准则。另一种是普遍主义的看法,认为虽然文化的进化和个人的修养存在着不同的历史和环境,但道德的基础仍然是具有普遍性的。对于这种普遍性的根源也有两种不同的说法,休谟认为普遍性的根源在于人类普遍的欲望,而康德则认为普遍性的根源在于人类的普遍理性。

上述种种分歧实际上源于一些看法的片面性,从系统论的全局立场看,产生这些分歧只是由于过分强调一方而忽视了另一方,当我们弄清了各种观点在整个系统中的作用和它们之间的关系,对立的观点是可以统一在一个理论体系之中的。例如主张追求个人幸福的个人主义和主张全局功利的全体主义之间,存在着不可分离的相互支持关系。个人主义是全体主义的基础,全局功利只能是个人利益(眼前利益和长远利益)的总和,离开了个人利益这一基础,全局功利是无法进行评价和计算的。反之,正是由于合作比单干能够获得更大的个人利益,人才有可能集聚成群体,组织成社会,因此社会是人的个人利益最大化的产物,全局功利的最大也就意味着个人利益的最大,完全无视全局利益会极度损害个人利益。

在伦理学的相对主义和普遍主义的分歧之中,包含着与一元论和多元论之间的分歧相类似的因素。从文化进化的角度看,各种文化所经历的不同进化历史导致不同的伦理道德体系是很自然的,正如不同进化历史造就了千变万化的物种一样。但是伦理学要想成为一门有理论体系的学问,就不能满足于相对主义,即使对于

千变万化的生物种来说,也存在着达尔文适应度这样明确而普遍的共同基础。文化的基础是属于同一物种的人,具有相同的生物价值体系,因而能够具有更为明显的普遍性。至于普遍性的根源,欲望和理性这两种说法都是不够深刻的,理性是在欲望指导下产生的工具,而欲望又是生物价值观的表现。

由此我们可以看出,要想为伦理学设定一个非形而上学的基础(出发点),只有选择人的生物价值体系作为道德产生的根源。比起其他基础,人的生物价值体系是在生物几十亿年以及人类几百万年的进化史中形成的,在以千年为单位的文化变革过程中是充分稳定的。又由于人类属于同一生物种,所以其生物价值体系具有普遍的一致性。因此把人的生物价值体系作为伦理学的基础,在时间和空间两方面都具有充分的稳定性。如果把人的本性看成“一半是天使,一半是野兽”的话,那么上述的观点就是主张从野兽出发去建立天使的特性,即天使是源于野兽的。从这个基础出发我们可以得到的一个结论就是“人之初,性本善”,在主张“人性本善”的性善说和主张“人性本恶”的性恶说之间,我们有充分的理由选择前者。这个理由是非常简单的,人的生物价值体系是经过了几十亿年的生物进化和几百万年的人类进化的千锤百炼得到的,相信它是迄今为止的大成至善的体系是完全合理的。我们目前还没有掌握一种改变人类生物价值体系的方法,即使到了将来我们有能力运用基因工程的手法来改变人的生物价值体系的时候,也是需要极为谨慎从事的。生物价值体系最深层的基础是达尔文的适应度,在这个基础上可以建立起善理的三大原则——存在原则、经济原则和预测原则。

主张“人性本恶”的人常常会列举一些人类生物本能中“恶”的因素作为性恶说的论据。但是把这些因素规定为“恶”却仅仅是依据一些关于“善”的形而上学的定义而已,未必有什么充足的根据。例如人们可能认为“懒惰”与“贪婪”是人性中“恶”的一面,但是懒

惰与贪婪的正确结合却正是经济原则的体现,即消耗较小的代价而获得较多的收益,如果没有这个重要的指针,我们连正确的走路方式也学不会,我们的一切决策都将失去依据。我们关心的是收益与支付之比,而不是单纯的减少支付或增加收益。懒到连套在脖子上的大饼也不转而饿死,或是贪到触犯了刑律都不仅不合乎经济原则,也违反更为基本的存在原则。某些宗教教义把性欲甚至食欲也看成是一种“恶”的表现。显而易见,这二者欠缺了任何一方都会导致人类的灭亡,违反存在原则。但是,关于人性本善的论证显然不应该使我们得出这样一条结论:所有按照自身欲望去作的事都是符合道德的。这种过于简单化的结论不仅不符合社会现实,而且使得伦理学本身成了不必要的东西,从根本上取消了所有关于道德的问题。

那么,本善的人性为什么需要受到道德的约束呢?这是主张性善说的人所必须回答的一个问题。(主张性恶说的人面临的是一个更困难的问题:性恶的人怎么可能创造出道德观念来呢?这个问题不求助于上帝的力量是很难找到答案的。)首先,一个不与其他任何人发生任何关系的人是不存在任何道德问题的。一个生活在孤岛上与世隔绝的人可以听任欲望的指挥,做他想做的任何事情。只有在人们由于合作的利益而聚集成群体,组合成社会的情况下,才会产生道德的需求。对于这种需求可以分为两类,一类是仅仅源于集群而与社会结构无关的道德需求,我们称之为“非结构性道德”需求;另一类是由社会组织结构所要求的道德,我们称之为“结构性道德”需求。非结构性道德由于只源于人与人之间关系,与社会的结构形式无关,所以不因社会结构的不同而不同,也不随社会结构的变化而变化。只要存在人与人之间的关系,就存在类似的非结构性道德要求,因此非结构性道德是具有普遍性的,在时间与空间方面都很稳定。反之,结构性道德是某种特定的社会结构所要求的,因此不同的社会结构可能有不同的道德需求,社会结构的变化

也会导致道德需求的变化,结构性道德具有较为明显的相对性。

道德需求产生于在集群和社会中只凭个人追求最大利益的生物价值观所决定的欲望行事反而不能通过合作获得个人的最大利益的情况。囚犯困境就是这种情况的一个典型的数学模型。囚犯困境问题所规定的前题条件不涉及社会的组织结构,因而是一个非结构性道德需求的例子。这个例子的阿克塞洛德解显示了在熟人的经常性交往之中道德是可以自发产生的;而哈丁解则表示在彼此不认识,不经常交往的情况下道德需要外界的强制力量。囚犯困境的阿克塞洛德解证明“礼尚往来”,“一报还一报”是进化稳定战略。从这个结论中我们很容易推导出通常的道德黄金律:“己所不欲,勿施于人”。因为在一个大多数人按照一报还一报的原则行事的群体之中,对他人的伤害最终会反射到自身,构成对自己的伤害。由于人不是先进化成人再聚成群体,而是从群居的类人猿进化而来的,所以非结构性道德会具有一定的遗传基础,存在于生物价值体系之中。“恻隐之心,人皆有之”就说明了这一点。

囚犯困境问题的阿克塞洛德解的重要意义在于从生物学前题出发,通过严密的数学方法导出道德的起源及自发产生的过程,从而提供了一个从非形而上学的基础出发建立伦理学理论体系的范例。这个范例所代表的精神实质就是通过利益分析判断道德的价值可以作为伦理学的基本方法,从而可以使伦理学摆脱形而上学的出发点,进入科学的领域。当然,囚犯困境问题是一个经过高度简化和抽象化的问题,实际的道德问题往往复杂得多,但是利益分析的方法也完全可以用于更复杂的情况。例如在囚犯困境中假定每次对战的利益或损失都是一样的,但现实之中则往往是不同的,因此“一报还一报”的简单报复主义有时会导致系统的不稳定。由于人对于他人所受损害的感觉不如自身所受损害的感觉那样直接,因而会造成报复偏高的倾向,如果双方报复偏高系数的乘积大于1的话,就很容易由正反馈而产生“冤冤相报何时了”的状态,使

双方利益都受到很大损失。小到日常争吵的逐步升级,大到村落之间的“打冤家”,这种正反馈不稳定性实例是很多的。即使在囚犯困境这种简化条件下,由于报复系数等于1,仍可能产生不衰减的报复循环,例如在一报还一报的战略中如果以很低的概率偶而背叛一次,那么在背叛之后就会产生很长的二者交替合作和背叛的循环,产生利益损失。因此从减少报复成本所带来的利益,就可以计算出“冤仇宜解不宜结”这种忍让的道德准则的价值。运用利益分析的方法研究道德的价值,使我们能够建立起更具科学性的伦理学——计算伦理学。有了这种定量的、可计算的伦理学,我们就很容易回答反道德分子经常提出的刁难:“道德值多少钱一斤?”

阿克塞洛德解的前题是熟人的长期交往,因此在一个人员流动很少的小村落里很容易自发地产生和维持互助的淳朴民风。随着社会规模的增大,流动性增加,人需要经常和不认识的人打交道,阿克塞洛德解的前题不再成立,此时就导致了哈丁解,即大社会需要外在的权威来实现最佳化,这种需求使得道德的权威性提高,逐步神圣化和形而上学化。“江湖义气”也就是在人口流动的背景下产生的。另一方面,社会规模的增大必然导致结构的产生和复杂化,从而产生了结构性道德需求。由于社会结构是文化进化的产物,变化较快,一种社会结构最长也不过几千年历史,因而结构性道德所要求的价值观不太可能通过进化过程影响生物价值体系,所以只能依靠外部强制的方式维持,从而促进道德的形而上学化的倾向。

道德形而上学化的倾向是这样产生的:对于非结构性道德来说,哈丁解的存在使道德需要外在的强制性;对于结构性道德来说,由于不存在与之对应的生物价值观,所以也只能以外部权威的形式存在。这种外部权威的强制力必需充分强大,以至于足以和本能的生物价值体系抗衡,否则就不能起到道德的作用。而且在文化进化过程中产生某种需要的道德时,人们往往还不知道它存在的

充分理由，正如一切生物进化都不需要理性的指导而能自发产生一样，在人们还不懂得遗传学时，就早已有了避免近亲繁殖的伦理观。即使在人们懂得了道德的可计算的价值之后，这种复杂的计算也不是每个人都能掌握的，而道德又要在绝大多数人受其约束时才有意义，所以迷信仍然需要，只是从对宗教的迷信转变到对科学的迷信而已。出于对道德的非自发的外部强制性需要，以及理论上普遍存在的无知，道德必然以一种不容批判的至高无上的面目出现，或者是宗教教义，即神的启示，或者是被神圣化了的圣人、伟人的教导。

道德的神圣化及形而上学化在道德实践之中是有效的，甚至是必要的，但对于道德理论即伦理学来说却是无用的，甚至是有害的。如果在理论上认为某种道德准则是神圣的，是形而上学地不受批判、不可分析、不可计算的，那么伦理学就不可能成为一种理论或学问，而只能是宗教戒律或圣人语录的集合，充其量再加上对这些条文的解释，如同元素周期律出现之前的化学，或是进化论之前的生物学。从文化进化的角度看，某些宗教教义或圣人之言能够在几千年的文化进化之中不被淘汰，似乎可以成这些准则合理性的证明，这就是我们通常总是倾向于尊重传统的原因。但是进化过程中可能出现的“逃离效应”使得进化未必总是得到最佳的选择。在生物进化之中，逃离效应产生于选择与被选择之间的遗传相关，由于这种遗传相关并不是普遍的规律，所以在生物进化之中逃离效应并不多见。而在文化进化之中，却存在着很多促成逃离效应的因素，例如道德的神圣化和形而上学化，使其处于不受批判、不可分析的地位，从而避免选择的竞争；道德的基本价值观是社会学习的前题，通常是在幼儿时代就被灌输了，人们往往会对这些基本价值观产生和生物价值体系相类似的感觉，在没有遇到文化冲突时，根本想象不到对其进行研究和检验，即使在文化冲突之中，人们也经常习惯于拒绝对方的文化，而很难

审察自己的立足点；此外，宗教教义和圣人言论通常是一个体系或集合，其中包含了正价值的成分和负价值的成分，只要总的正价值大于总的负价值，这个体系或集合作为整体就不会被淘汰，从而使具有负价值的有害的道德准则作为整体的一部分而得以保存。造成逃离效应的另一个原因与生物进化中相似，即道德的社会性。道德准则只有在被社会普遍接受时才有意义，因此当某一个道德体系被社会普遍接受时，另一个道德体系即使更为优越，在被普遍接受之后能带来更大的利益，当只有少数先驱者执行这一道德准则时也只能带来损失，如同在公共汽车站上多数人乱挤时少数人排队是没好处的。这就象多数孔雀选择长尾时，即使短尾更利于生存，个别孔雀选择短尾也会降低其繁殖适应度。“入乡随俗”就是基于这样的理由。因此一种文化价值观体系一旦被一个群体普遍接受，就会具有顽强的自我保护能力，这种能力一方面带来了文化的稳定性，一方面又有抵抗优化，造成逃离效应的作用。

逃离效应造成的文化非最佳化的可能性，使我们可以得到非形而上学的伦理学的一条基本原则：任何宗教教义、圣人之言、多年的传统习惯，都不是当然正确的，都不能作为伦理学的出发点，都不能回避利益分析的解剖刀，只有对社会和组成社会的人带来利益的道德，才是可接受的。与此有关的另一条原则，我们称之为“道德体系的最小化原则”。这个原则所主张的是，一个好的道德体系应该只包含经过利益分析证明是对社会有利的带有正价值的道德，而不应包含带有负价值或正价值不大的道德。就是说，一个好的道德体系应该是经过最小化的道德体系，其中只包含必要最小限度的道德准则。或者说，我们所接受的道德准则应该是越少越好，而不是越多越好。这是由于接受任何一种道德准则都不是无代价的，任何一个道德准则都意味着在人的决策过程中对人的生物价值体系的干扰，使其偏离最佳位置，形成一定的损失，称

为道德成本。如果道德不起这样的作用，而和生物价值体系完全一致，则我们没有必要把它算成一种道德，例如我们不会规定饿了应该吃饭，困了应该睡觉是一种道德。因此如果一种道德带来的利益很小，不超过其成本，那么这种道德就没有理由被接受。道德成本不仅包括价值观偏离损失，还包括道德风险。所谓道德风险是指这样一种情况，即任何一种道德所带来的利益都是有限的，道德的合理作用应该是对生物价值体系造成适度偏离，使之在社会环境下实现最佳化。但是人们在实践中对于道德的利益未必总有恰当的估计，加之道德本身的形而上学化倾向会使其走向绝对化，产生道德至上的偏差。由逃离现象使总价值体系形成过度偏离，造成价值损失的可能性，构成道德风险。欧洲中世纪的宗教审判，中国吃人的封建礼教，都可以显示道德风险有时是非常大的。就从减小道德风险的角度看，对道德的价值作出正确评价也是必要的。

当社会结构发生转型时，很多结构性道德需求由于原有社会结构的消失而消失，所对应的结构性道德也因而失去了价值，不再受人们重视，这是社会自发的道德体系最小化过程，是合理而必然的。形而上学道德家们不懂得道德体系最小化原则，把这一现象看成是“道德沦丧”、“道德滑坡”而又无可奈何。实际上这个最小化过程正是道德体系进化、适应和优化的过程。除了这种时间轴上的道德变革之外，在空间上也可以看到道德体系最小化原则的例子。我们可以想象一个“留学生社会”，在一个留学生宿舍中聚集着一群来自各个不同国家，具有不同文化背景的留学生，任何一种文化都不占主导地位，在他们之间也没有必要按所在国的风俗习惯交往。在这样的小社会中也会形成一种文化，产生内部的道德体系。显然，这里能够为全员普遍接受的道德体系只能是各种文化中道德体系的公约数，而不可能是公倍数。也就是说在多种文化平等融合的时候，只有各文化中都存在的道德准则

可以被普遍接受，各个文化中特异的道德准则将被拒绝，所有文化中的所有道德准则都被普遍接受的情况是不可能发生的。因此在文化融合时道德体系的重构也是遵循道德体系的最小化原则，而不是最大化原则。当我们对某种道德准则的价值认识不清时，最小化原则可以作为重要的指针，这种道德准则如果是其他文化的道德体系中所不包含的，那么其价值是特别值得怀疑的；而所有文化中都包含的道德准则，其具有正价值的可能性较大。

从利益分析确定道德价值的方法表现了崇尚理性和反对形而上学的基本立场。理性并不能使我们获得真理，但是理性可以帮助我们寻找最佳点，克服形而上学所造成的逃离效应。与非理性的形而上学相比，理性具有自我修正和自我优化的能力，当我们所依据的理论出现偏差时，可以按理性的原则对理论进行修正，使其进一步优化，而非理性的形而上学则缺乏这种功能，可能使我们长期处于非最佳状态。用利益分析来理解道德价值的另一方面的优越性在于，当道德原则与实际利益发生冲突，或是两种不同的道德原则之间发生冲突时，我们可以按综合利益（包括长远和全局利益）最大化的原则通过取舍、权衡和分离等各种方法寻求最佳的策略，如同囚犯困境中根据对战继续的概率来计算应采取合作还是背叛的战略那样，而不会陷入道德与利益、道德与道德之间的两难处境而不知所从。这种方法为道德和利益设定了共同的坐标，提供了对二者进行综合决策的统一方法。应该看到的是，大多数正常人的头脑中进行决策时也是按相同算法进行的，只是各人对各种道德准则的评价价值有所不同，因此这种方法也有助于对他人道德行为的预测。从囚犯困境的例子可以看出，道德的利益分析和价值判断是比较复杂的，有时需要相当高度的数学技巧，以下我们用一些实际例子来说明怎样分析道德的价值和某种道德可以被接受的理由。

我们先用一个简化的伦理学问题作为例子。在这个伦理学问

题中有两个医生，第一个医生收治了 6 个病人，他们各自有不同的内脏患病，只有通过脏器移植才能治好，医生把其中 1 个病人的健康内脏分别移植给其他 5 个病人，治好了这 5 个病人而杀死了作为内脏提供者的病人。第二个医生接收了一次交通事故的 6 个负伤者，都需要尽快手术，其中一个伤处较多，手术需要 1 小时，其他 5 人伤口少，每人手术只要 10 分钟，但每一个伤员都必需在 1 小时内作完手术，否则就会由于失血过多而死亡，转院或叫别的医生在时间上都不允许，这位医生先依次给 5 个伤口较少的伤员作了手术，伤口多的伤员则由于来不及抢救而死亡，但是如果先抢救伤口多的伤员，则其他 5 个人会死，所以医生选择了前者。在这个问题中的两个医生的行为结果都是牺牲了 1 个人而救了 5 个人，在表面的利益计算来看，二者的行为是同样符合道德的，但是从我们的道德直觉来看，第一个医生是明显违反道德的。虽然我们的道德直觉未必都是正确的（如果都正确就不需要伦理学了），但是可以帮助我们发现问题的所在。有人认为二者的区别在于加害于他人和拒绝救助他人道德上有很大的不同。这种说法不无道理，但是作为第二个医生如果在只有一个伤员的情况下拒绝救助，同样是在道德上不能容忍的行为。第一个医生的最大的错误在于他违反了职业道德。职业道德是一种结构性道德，职业道德的作用在于使该职业得以存在，因此违反了职业道德就违反了这种职业的存在原则。如果医生可以按照第一个医生的方针行事，那么医生作为一种职业是无法存在下去的，因为多数人甘愿冒百万分之一的失事概率的风险去乘坐飞机，但恐怕很少有人愿意冒六分之一被杀概率的风险去医院看病。职业道德一般都具有维持该职业存在的作用，例如作心理咨询的人不能泄漏患者的隐私，否则患者不来或不说实话，心理咨询就不可能存在了。

结构性道德的价值取决于它所维护的社会结构的价值和该道德在维护该结构中所起的作用。社会结构是文化进化过程中发明

的产物，任何一种社会结构都不是绝对的、永恒的，人们总是可以发明出新的社会结构取代旧的社会结构，因此社会结构的价值总是相对的，在只考虑一种社会结构的情况下，其价值可以由该结构与无结构状态的比较来确定，而在比较两种以上的社会结构时，其价值是由各结构之间的相对比较来决定。因此当一种新的社会结构出现时，旧社会结构的评价可能会大幅度下降，由原来与无结构状态比较所得正值，变为与新结构相比较所得的负值。与此同时，用于维护旧结构的道德价值也会随之下降。在对社会结构的评价之中，道德成本也是一个重要的因素，如果一个社会结构所带来的利益不大于为维护这种结构所需道德所支付的道德成本，那么这个社会结构本身就是不合理的。同样，如果结构 A 与结构 B 对社会起到相同的作用，而维护结构 A 所支付的道德成本低于结构 B，那么结构 A 的价值大于结构 B，用结构 A 取代结构 B 是合理的。

无论是结构性的还是非结构性的，对道德的评价都不能离开综合权衡的系统论的观点。例如对于利己主义和利他主义的评价，争论的错误往往来源于把二者看成是完全对立的观念。实际上，离开利己的利他主义和离开利他的利己主义都是不可能存在的。绝对不利他的纯粹利己主义的例子是杨朱之道“拔一毛而利天下，不为也”，作为一种文化，这种哲学是自我毁灭的，因为它不可能被传播开，著书立说总是为别人看的，传播这种思想本身就是利他的行为，与这种思想是矛盾的。从实践上看看，任何社会的分工合作之所以能够成立，都包含利他的因素，完全拒绝利他，意味着从社会中退出，这样的人当然得不到社会合作所带来的利益，结果是违反利己主义的，而对于孤立的个人来说，利己主义作为一种道德准则是没有意义的。同样，完全排除利己的纯粹利他主义也必然走向悖论。如果社会中有一部分只讲奉献不讲索取的纯粹利他主义者，那么社会将形成不公正的剥削，而使得利他主义被

淘汰；如果社会全部由纯粹利他主义者组成，那么由于无人讲索取，社会没有任何根据说明应该奉献什么，社会将会失去行为目的，变成一个巨大的垃圾堆。只有通过利他的手段达到利己的目的，社会才有存在的根据，才能够具备优化机制所必需的评价准则。

价值分析的方法与形而上学的方法相比，更容易获得结论而避免无意义的争论。例如对于杀人为什么是一种不道德的罪行的解释，利益分析和形而上学会有不同的出发点。形而上学的解释需要事先设定一个形而上学的出发点，例如生命的神圣尊严之类，而利益分析则是从任何人都不愿意自己被杀死的事实出发，这一事实是符合存在原则的。如果没有不杀人这种道德规范，一个社会是不能成立的，因为当合作所带来的利益不足以补偿很高的被杀概率所带来的风险损失时，人们所能选择的最佳决策是退出这个社会。表面看起来，形而上学的方法和利益分析的方法在这个问题上得到的结论是相同的，但是推论下去就会有不同的结果。从生命的神圣尊严出发，可以推论出杀狗吃肉也是不道德的，但是利益分析则得不到这种结论。把杀狗吃肉规定为不道德的行为会导致理论上的矛盾，即为什么杀狗吃肉是不道德的而杀牛或杀猪吃肉就不违反道德呢？以生命的神圣尊严为前题来讨论堕胎是否违反道德，则会遇到一个棘手的问题：胎儿是不是生命，胎儿从什么时间开始算人？而利益分析的方法就不会遇到这个问题。如果生活在一个谁都可以随便杀人的社会里，每个人都会感到恐怖而惶惶然不可终日，但是生活在一个可以自由堕胎的社会里，没有人会感到这种恐怖，因为凡是能够感到恐怖的人，都已经没有被堕胎的危险了。按照道德体系的最小化原则，把堕胎看成不道德的行为显然是没有意义的。

从形而上学的观点来看，理想主义与现实主义是对立的两种主张；但是按照系统综合的利益权衡的观点来看，理想主义与现

现实主义的关系就象全体主义和个人主义、利他主义与利己主义的关系一样，是相互不可分离的互补关系。在一个人或一个社会的意识形态之中，理想主义与现实主义恰成软硬结构的关系。现实主义追求个人或社会的最大利益，是小时间尺度的优化机制，是保证个人或社会得以存在和经济地运行的硬结构；而理想主义则是超越现实的作用力，属于高风险而可能产生高收益的软结构。从地形图模型来看，现实主义的作用是趋向最佳点，而理想主义则是造成对最佳点的偏离，必然产生一定的损失，这就是其风险的所在，但是在一个复杂的地形图之中，对局部极小点产生偏离的逃离效应有时可以起到跨越山口找到更深的极小点的作用，此时就会产生巨大的利益，但是获得这种利益的概率是很低的，所以风险是很大的。因此对于一个人或一个社会来说，缺少现实主义则会威胁其存在，缺少理想主义则会妨碍其发展，二者需要一种平衡。例如一个国家的科技政策中，对于基础研究和应用研究需要有一个适当的比例，基础研究相当于理想主义的成分，投资风险很大，可能一时看不到效果，但是取消这种投资则会妨碍科学技术的长远发展；应用研究属于现实主义的部分，必须占有一定的投资比例，否则会失去现实的利益。对于一个企业来说，也存在着用于研究开发和维持日常生产的投资比例问题。对于一个人来说，也存在理想主义和现实主义的平衡问题。“人是要有一点精神的”，这里的精神指的是理想主义成分，但是“一点”则表示不可过量。完全没有理想主义，会使人成为蝇营狗苟之徒，失去大发展的机会；而过度的理想主义则会带来过大的风险，现代诗人的自杀概率较高，就是理想主义风险的表现。“生命诚可贵，爱情价更高”这样的诗句就表现了明显偏离最佳点的价值观体系。

中国的传统道德准则之中，包括了忠、孝、节、悌、礼、义、廉、耻等典型概念。忠孝节悌是典型的结构性道德，所强调的是社会地位的序列和由序列所决定的服从关系。由于中国在历史上

多数时期的社会结构是由治水需要而产生的高度中央集权的大社会组织，维持这种大规模社会组织需要多层次的树结构，而树结构能够成立的基本前题就是下级对上级的服从关系。“忠”就是对国家机构或其代表者君主的绝对服从，作为组织的一个成员努力维护，至少是不破坏组织的利益。没有这样一种准则，任何组织都会被瓦解，但是组织越大，这种道德需求就越大，因为组织越大，层次就越多，信息传递失真的可能性就越大，同时成员越多，出现内部破坏者的概率也越大。在君主制的结构被其他结构取代之后，作为国家机器的树结构依然存在，社会中的大规模组织（例如大企业）内部也需要多层次树结构，因此忠于职守，忠于某一组织的道德需求仍然存在。在以市场经济为代表的无上下级层次的网状关系之中，例如企业之间和个人之间的平等关系之中，由于不存在服从关系，“忠”也就不再是一个重要的道德准则，而代之以基于预测原则的“信”。在任何合作之中，信用不足意味着环境的可预测性下降，整个社会运行的风险成本增加，对社会中所有人都是不利的。诚实很多宗教教义之中都被规定为一种形而上学的道德准则，但是诚实本身并没有什么道德价值。从“狼来了”的故事可以看出，诚实只有在涉及信用时才由于信用而产生价值。在实际生活中，“善意的谎言”或无害的谎言是很多的，因为并不是所有的谎言都会造成信用损失。

与“忠”相比，孝、节、悌虽然也是一种结构性道德需求，但其普遍性比“忠”要差得多。这三个概念实际上是把“忠”这种个人与组织的关系向着父子、夫妻、兄弟关系所作的推广，但这种推广只在理论的对称性上有一点美学意义，而缺乏现实的结构需求。在一个以经验为主要知识来源的农业社会之中，作为生产基层组织的四世同堂式的大家族，由父系血缘关系所规定的树状结构要求“孝”所规定的小辈对长辈的服从。既使在这样的社会中，孝与忠也只是形式上的相似，并没有明显相互支持的作用，

“忠孝不能两全”的矛盾局面时有发生。在一个生产和教育充分社会化的社会，对于核心家族来说对“孝”的结构性道德需求基本上消失。在中国当代社会中对“孝”的理解主要是指子女对父母的赡养义务，但是在现实之中对“老有所养”的解决方案里，“孝”既不是唯一的也不是最佳的方案，社会化的养老保险比“孝”要可靠得多。因为并不是每个人都有子女，也并不是每个人的子女都有充足的能力赡养父母，更不是每个有能力的子女都乐于赡养父母。母爱是提高达尔文适应度的因素，在很多动物特别是哺乳动物中普遍存在于基因之中，但是赡养父母对于动物的适应度毫无贡献，所以不可能在进化中产生。因此“孝”不存在于人的生物价值体系之中，“孝”作为一种道德也就具有较高的道德成本，维持这种道德也就比较困难，“痴心父母古来多，孝顺子孙谁见了”也就是一种普遍现象。在“孝”的概念之中也包含与生物价值观一致的成分，那就是“不孝有三，无后为大”，体现了一种生殖欲望，与提高适应度的目标一致。但是从全社会的角度看，这种生殖欲望的结局却是使全社会陷于马尔萨斯状态，为了脱出马尔萨斯状态，有必要从哈丁解出发，采取外部强制手段抑制这种生殖欲望。综上所述，“孝”作为一种道德是利益甚小而成本颇高的，是一种将被淘汰的道德。同时，反对避孕的宗教教义也是一种明显的“自然主义谬误”，驳斥自然主义谬误是很容易的，穿衣服、住房子都是“不自然”的行为。

在父系血缘的大家族中，“节”所规定的妻子对丈夫的服从具有一定的维护树结构的作用，但是“节”与“忠”也只是形式上的相似而缺乏内在联系，一个惧内的官员未必不是个忠臣。“节”所包含的另一种意义是贞操观念，贞操概念的一种意义是指不变性，这一点可以用与可预性有关的道德观念所取代，例如“忠”和“信”等等。贞操的另一个含意是带有专一性的性道德，这是一种结构性道德，其根据是父系血缘关系所确定的男性对下一代的养

育义务，以及性与生育之间的必然联系。在避孕手段高度普及的情况下，性与生育之间的必然联系被切断，世间的贞操观念也会逐步淡漠。因此“节”也是一种高成本低利益的道德准则，在文化进化之中会被逐步淘汰。“悌”在父系大家族的结构中所起的作用也不大。只有在父亲不在时才起到一点决定顺序的作用，在核心家庭之中则几乎没有作用，因此“悌”是个不常被强调的道德，而且除了儒家文化圈之外的文化中很少有其对应概念。如果给一个美国儿童讲孔融让梨的故事，他一定大喊“这不公平！”在儒家文化圈之中，“悌”也往往只在推广的意义上起作用，如同“桃园结义”中映射的兄弟关系只是一种顺位的代名词，华人和日本人的黑社会组织都把上级称为“大哥”就是用兄弟关系来表现顺位，是“悌”的推广应用。而在欧美的黑社会就不这样称呼，而称黑社会头目为教父，就是因为在欧美的文化中没有“悌”这个概念。“悌”是一种已经消亡了的道德准则。

“礼义廉耻”基本上属于非结构性的道德，或者是对具体结构依存性不强的通用结构性道德，因此具有较高的普遍性。“礼”是人与人之间通过某种语言、行为、仪式等符号信息交换来确认其相互关系的手段。“非礼”则指不符合相互关系的行为。不论是结构性的还是非结构性的人际关系，都需要“礼”作为规定的符号进行信息交换，因此“礼”是具有普遍性的。但是由于符号编码的自由度很高，“礼”的具体内容和形式可以由于文化、结构、时代的不同而千差万别，所以“礼”的形式又具有特异性。当少数人进入一种多数人的文化环境时，“入乡随俗”是通常的手段。在多种文化的平等融合之中，“礼”往往趋向于耗能较少的简化形式。在非结构性的人际关系之中，“礼”主要用来确认礼尚往来的原则，即“来而不往非礼也”，这恰好是囚犯困境的阿克塞洛德解。

“义”的含意比较广泛，通常是指对抗邪恶、帮助弱者，或不损害他人的道德准则。这些道德准则都是非结构性的。因此在不

同的文化背景和不同时代具有普遍意义。在现代的大规模的发达社会之中,重要的道德准则需要用法律的形式保证其外部强制性,法律与道德是相互支持的,如果没有对某种道德的普遍承认,与之相关的法律会由于法不责众而成为一纸空文,反之没有法律的强制性约束,被多数人接受的道德也会在哈丁解条件下被少数人破坏,如同公共汽车站的乱挤现象。在大规模社会之中,对抗邪恶的法律手段比路见不平拔刀相助的侠客更有力,帮助弱者的税收和福利政策比仗义疏财的个人更可靠。没有完备的民法,不可能有效地防止损害他人的行为。

“廉”是一种结构性的道德,但由于它是所有结构都需要的,因此具有普遍性。“廉”的基本原则在于个人作为组织结构中的一个功能元素时不应出于自身利益而破坏应有的功能。如果这条原则得不到贯彻,社会中的任何组织结构都将会被破坏。由于人的生物价值体系是追求个体利益的,所以“廉”的道德成本比较高,但是“廉”又是维持结构功能所必需的,因此其价值也很大,这也就是“高薪养廉”政策的理论根据。当然,如果某种结构不足以造成超过其运行成本(包括道德成本)的社会效益,那么这种结构本身就失去了存在的理由。

“耻”是一种非结构性的道德,因此在任何文化背景下的任何时代,“无耻”都是一个否定性的词汇。“耻”是指对群体中其他个体对自身某种行为产生的或可能产生的评价作出相应反应的能力。对于任何集群的有高度学习能力的特别是具有文化传递能力的动物来说,这种对其他个体评价的反应能力都是有意义的。例如狗会对人的赞赏或指责的态度作出相应的反应,人们通常在进行决策时考虑“别人会怎么说”也是这种能力在思维过程中的表现。“知耻”的价值是很明显的,其进化过程类似于动物斗争仪式化的进化过程,仪式化的斗争对于斗争的双方都达到了斗争的目的而降低了斗争的成本。如果一个社会由无耻之徒组成,那么社

会对每个个体的行为制约都需要通过给个体造成实际损害来进行，社会的运行成本是非常高的。由于“知耻”的能力产生于生物进化过程，存在于人的生物价值体系之中，所以“耻”的道德成本是比较低的。有时在社会意识形态中可能存在着由于逃离效应而产生的道德偏见，那么认识到这一点人需要克服“耻”的约束向多数人的观念挑战，只好“走自己的路，让别人去说吧”，非此社会是不能进步的，只是我行我素者通常要承担开拓的风险。

关于西方文化中的一些道德观念，可以举出“自由、民主、人权”等作为例子，用利益分析的方法评论其价值。“自由”初看起来是个令人难以理解的道德观念。如果自由意味着为所欲为，那么事实上就是人按照生物价值体系的指导来行动，而道德的需求恰恰是社会的利益需要对生物价值体系进行约束和修饰而产生的，自由作为一种道德正好与其他道德相对立。如果把自由看成是按照道德观念行事的自由，那么自由仅仅是生物价值体系和道德修饰的综合效果，作为一个独立概念没有太大的意义。这个问题涉及到东西方文化的差异。在中国历史上，文化转形往往伴随着道德体系的最小化过程，从孔子时代的“礼乐崩坏”到五四运动的打倒孔家店，乃至当代的“道德滑坡”都是对传统道德的否定，通过降低旧道德的评价来修正逃离效应所造成的对最佳点的偏离。在一个最小化的道德体系之中强调生物价值体系的“自由”是意义不大的。西方文化的道德体系多数立足于宗教教义，具有很强的形而上学性。这种形而上学的道德体系不按“权衡”的方式运转，而遵循“全或无”的原则。例如尼采宣布“上帝死了”的全面否定的态度，或是把旧的意识形态放在理性的审判台上，要么宣布存在的理由，要么被消灭。但是有很多道德准则都是不能被消灭的，而按照全或无的原则，只要不被消灭就依然保持其形而上学的威力。这种形而上学的道德体系很难进行适度调

整，克服逃离效应，于是人们就发明了“自由”、“人道主义”这样的概念，给人的生物价值体系加上一道形而上学的光环，使其有足够的威力去对抗其他形而上学的道德准则，以克服逃离效应。因此“自由”这一概念在西方文化的道德体系中起着重要的作用。这种方法虽然能起到和最小化道德体系类似的作用，但在理论上有很大缺陷，道德体系最小化原则可以给出各种道德的相对价值以用于决策，而多种形而上学的对抗则只给出类似于两个无限大相比较的问题，在实际决策中只能依靠直觉。

从哲学的高度看，自由有着重要的价值。通常的看法认为，自由是对必然的认识，只有认识了客观规律，人才能获得自由。如果把认识看成一个广义进化过程，那么自由与认识是软硬结构的关系，自由在软结构中对认识进行建构。进化的基本机制是变异和选择，而变异则来源于自由。因此，没有自由就没有变异，没有变异就没有进化，没有进化就没有认识的发展。在自由与认识的因果关系中，自由是因，认识是果。同时，认识在建构下的增长又支持更高层次的自由。

“民主”的内容主要包括三大原则，即多数原则、程序原则和少数原则。多数原则就是少数服从多数的原则，这个原则的根据是集团斗争的仪式化，在两部分人的斗争之中，力量与人数成正比，实际斗争的结局通常是多数得胜，用表决来代替实际的斗争，可以大大降低斗争成本，这就是多数原则的价值所在。但是，多数表决并不能保证决策的正确或利益最大，而且通常运行效率比较低。虽然每个人都具备对自身利益较为准确的生物评价体系，但是复杂的社会结构也会使得这种生物评价体系未必总能有效地进行决策，特别是大众传媒的存在使得多数人的想法容易受到操纵，使民主变成仅仅是形式。因此在实际的社会结构中绝对的民主和绝对的集权都是不存在的，即使是君主集权的结构之中，也存在着“集众思，广忠议”的机制；在号称民主的制度中，也广泛存

在着操纵舆论的现象，并且为个人的独断留下了充分的余地。因此，多数原则并不具有绝对的价值，而是视结构要求而具有不同的价值，例如在军事或准军事组织之中，集权的价值是大于民主的。

民主的程序原则是仪式化所必需的。动物斗争的仪式化总是以各自的实力作为后盾的，而维持人类集团斗争的仪式化则要困难得多，仪式化虽然使斗争双方及整个社会降低斗争成本，但当斗争所争夺的利益大于实战成本时，破坏程序原则转入真枪实弹斗争的可能性是很大的。因此为了维护程序原则需要权威性的外界强制力量，例如在君主立宪制中的君主地位，或是基于社会共同宗教信仰的形而上学意识形态。在形而上学力量比较薄弱的东方文化圈中，只有在社会充分富裕，内战带来的社会破坏使实战成本极高，超过集团争夺的利益时，才能自发地维持程序原则。当社会比较贫困，处于马尔萨斯状态，总有一部分人处于生死边缘时，实战成本很低，程序原则是很难维持的。此时社会只能在集权和内战之间选择，而集权的成本显然小于内战的成本，集权的产生仍然要依靠内战中的自然选择，或是其他力量无法竞争的军事实力。

民主的“少数原则”是指少数人持有和多数人不同意见的权利。少数原则的根据是多样化的价值，一个所有人意见都一致的社会是不可能进步的，正如基因中不包含变异的种群不可能进化一样。表面看来，少数原则与多数原则是矛盾的，实际上二者分别应用于不同的场合，多数原则适用于社会的硬结构，而少数原则适用于软结构，只有在非要所有人一致行动，否则社会会遭受重大损失的情况下，少数服从多数才是必要的。对于一些问题，分离主义可以成为有效的解决方案，此时只要遵循一条原则，即利益与风险同在，权利与义务同在的原则。这条原则的理由是人类学习机制中的责任局部化需要。例如对于关于脑死算不算死亡，从

脑死的人身上摘取脏器用于移植是否合乎医学伦理的问题，往往很难通过争论取得一致意见。这个问题并没有必要用表决的方法取得一致，按分离主义的方案，可以让每个人自行决定选择 A 派或 B 派，A 派表示承认脑死，在他自己进入脑死状态时同意其内脏用于移植，同时在需要时有权接受来自脑死者的脏器移植；B 派表示否认脑死，选择 B 派的人在脑死时不用于提供脏器，而他在需要时也无权接受来自于脑死者的脏器移植。同样，按照对婚外恋的赞成和反对也可以分为两派，每个人可以自行决定加入哪一派，只要总是由同派的人结成配偶，两派之间是不会发生利害冲突的。“一国两制”也可以算是分离主义方案的例子。

“人权”是一个内容比较复杂的概念，其主要作用是把“己所不欲，勿施于人”的原则用明确的法律形式加以固定，使个人权利或利益受到侵害时有得到保护的法律依据。具体权利的价值可以按照其对社会造成的利益进行评价。例如著作权保护作者的利益，可以促进文化事业的繁荣；专利权保护发明者的利益，可以促进发明的公开，有利于技术进步；肖像权有利于促进商业收入的公平分配；名誉权可以保障“耻”的社会机制的正常运行。在不同的文化背景中，各种权利的价值是不同的。在发明只是偶发事件的农业社会中，和在发明需要巨额投资并带来巨大利益的工业信息社会中，专利权的价值显然有很大差异。隐私权的价值也随文化背景的不同而有所不同，在一个非最小化的由大量多余的形而上学道德观念统治的文化中，人们的一些并不损害他人的行为有时也会与多余的道德发生冲突，此时就极需要隐私权来保护自己；但是在一个运用最小化道德体系的宽容社会中，在不损害他人范围内的各种行为都不会受到非议，隐私权的价值也就不太大了。在中国现行的法律中不包含无罪推定原则，这一点看起来和“人权”概念不符，但是如果不是形而上学的看待人权，而是采用利益分析的方法，则需要计算无罪推定原则的得失，无罪推

定原则可以减少冤案，但可能使犯罪增加，需要计算的是在一定的安全成本的前题下，怎样能使个人受冤案所害和受犯罪所害的概率之和达到最小，从计算的结果可以判断是否应该采用无罪推定原则，这种判断要比形而上学的方法合理得多。作为“人权”的一种推广，在西方文化中产生了维护动物权利的主张。如果从保护珍稀动物，维护生态平衡，保持物种多样性的角度出发，那么是符合利益分析的。如果形而上学地谈论动物的权利，则是毫无意义的。从利益分析的角度看，反对虐待动物的理由并不是基于动物的权利，而是由于虐待动物反映了行为者的心理健康有问题，如果出于经济上的需要，则对动物本身的痛苦不必过于关心。为了心理健康的考虑，可以采用孔子的解决方案，“君子远庖厨”，又能吃肉，又不会损害同情心。因此澳大利亚的动物保护组织对中国的养熊者从活熊身上取熊胆汁表示抗议，在中国人看来纯属多管闲事，中国人看到澳大利亚人给羊注射药品使羊毛脱落以代替剪羊毛，也不必去想羊被注射后会不会感到不舒服。即使这种药品会使羊非常难受，我们也会认为注射是为了节约剪羊毛的工时以获得利益，而不是为了满足虐待狂的心理需求。

从比较文化学的角度看，东西方的道德体系有着深刻的区别。从最基本的道德黄金律也可以看出两种文化的深层精神结构有重要的不同点。中国的道德黄金律的表现是“己所不欲，勿施于人”，而西方的道德黄金律则是“己之所欲，施之于人”。在有些人看来，这只是同一种观点的不同表现而已，仔细观察可知这两种表现在理论上和实践上都有重大的差异。对于一个人来说，“所欲”是“所不欲”的补集合；但对于多数人的场合，这个关系不成立，“所欲”与“所不欲”是明显非对称的，就象“证实”与“证伪”之间的非对称性一样。“所不欲”是指一些引起不快的感觉的东西，对人来说是起保护作用的必要最小限的感觉，属于硬结构，因此在不同的人和文化之间一致性是很高的。疼痛、

疲劳、贫困、侮辱、饥饿，这些带负评价的事物对任何人来说都是“所不欲”的。“所欲”则属于软结构而含有大量变异，对于金钱、名望、权力、知识、宗教信仰等追求，每个人都有不同的侧重。因此一个不爱吃臭豆腐的人每次请客都不摆臭豆腐是没有问题的，而一个爱吃臭豆腐的人每次请客都摆上一盘臭豆腐，就会令人不快了。

“己所不欲，勿施于人”的原则看起来有点消极，但却符合道德体系的最小化原则；而“己之所欲，施之于人”则具有很强的强加于人的味道。关于中国文化中这种道德体系最小化的非形而上学倾向的起源，是个非常耐人寻味的问题；对此我们只能提出一个假说。象汉民族这样具有 10 亿人口以上的大民族，在世界上是独一无二的，因而可以想象，汉民族可能并不是一个单一的民族，而是一个众多民族融合的产物。这种民族融合可能是由治水中统一行动的需求产生的，那么这个融合过程可能发生于诺亚在方舟中漂泊的时代。由于多民族的平等融合，各种文化的求同存异经历了类似于“留学生社会”的建立过程，创立了道德体系最小化的非形而上学的思想传统。以上的假说目前还缺乏史实的根据，只算是一种猜测，但是汉族文化高度的融合能力却是得到历史证实的，其中最近的一次大规模的文化融合就发生在 300 年以内，当满族入主北京，成为中国的统治者时，拥有自己的语言、文字，短短 300 年之后的今天，满族只留下户籍上的意义，会说满语、能识满文的人已是凤毛麟角，在我们周围广为人知的满语只剩下“萨其马”这一个单词。

由融合产生而具有融合能力的中国文化，其最重要的特征及优点就是其非形而上学性和由此产生的实用态度。由于这种非形而上学性，中国文化中可以容纳各种宗教，但中国人对宗教的态度始终是实用的，工具性的。在中国的历史上从未发生过政教合一的现象，也从未有过基于意识形态冲突的宗教战争。在农民起

义之中，宗教往往被用作宣传和组织群众的工具，如同一面旗帜，人们利用旗帜作为集聚的标识，而旗帜上写的是什么字并不重要。太平天国是以基督教作为旗帜的，但是当造反失败，集聚的农民回到自己家里后，很少有人继续对基督教感兴趣。正因为宗教具有聚众作用，中国的历代统治者都对宗教持有一定的戒心。关于中国人对宗教的实用态度，毛泽东同志看得非常清楚，在他与斯诺的谈话中曾提到，中国人对神的态度往往是生了病信神，病好了就不信了，没孩子时信神，有了孩子就不信了。这种实用态度可以用一句中国的俗语来概括：“管他黑猫白猫，逮着耗子就是好猫。”但是实用态度不应与实用主义相混同，实用主义仍是一种主义，是一种形而上学。实用主义的口号是“实用即真理”，而实用态度所关心的是“好猫”与“坏猫”，不是“真猫”与“假猫”，如果一个机器猫能抓老鼠，也仍然是一个好猫。当世界末日即将到来的时候，西方人会将他们的钱交给教会，以求灵魂得救，而中国人则用这些钱买些平时舍不得吃的东西，暴撮一顿，以偿平生宿愿。在中国的经济转换期有人惊乎青年人的“信仰危机”，但这种担心是完全没必要的，在一个不以形而上学作为精神支柱的文化传统之中，没有信仰不会造成任何精神危机，中国的大街上也没有出现嘻皮士。

在通信交通高度发达的今天和未来，不同国家之间的经济联系日益密切，不同文化之间的交流日益增加，世界变成了小小的地球村。在这种形势下，避免文化冲突成为重要的课题，非形而上学的融合型文化将日益显示出其生命力。在多种文化的共存、交流乃至融合之中，“己所不欲，勿施于人”的原则比“己之所欲，施之于人”的原则更不容易引起冲突。形而上学的思想体系总认为自己是唯一的真理，有责任推广于全球。西方人进行文化扩散的先头部队是传教士，到处去推广他们的信仰，建设教堂，这种方法对于本身文化根基薄弱的民族可以取胜，对于持有另一种形

而上学的民族则是造成激烈的冲突，而对于象中国这样的具有深厚的非形而上学文化基础的文化来说则是无效的。任何一种形而上学在中国都会被转化为一种工具，而失去它的本来意义。中国人进行文化扩散则完全是非形而上学的，中国人所到之处是开设中国饭馆，不想吃的人决不强迫，听凭个人的选择，这种方法是不会引起文化冲突的。从历史上看，中华文明能历 5000 年而不灭亡，是与其非形而上学性密切相关的。古代很多曾经高度发达的文明最终走向灭绝，其原因很可能是形而上学所引起的逃离效应造成的。非形而上学的思想体系在文化进化之中不容易产生逃离效应，也就不会过度偏离最佳点，因而具有极强的生命力。

以上对各种道德和价值观的利益分析的例子，由于缺乏必要的事实根据和充分的数据，只能达到定性的水平，也含有猜测的成分，有些结论可能是武断的，甚至是充满偏见的。但是本节的目的并不在于对具体的道德作出评价，而是通过这些例子展示利益分析的方法，为伦理学奠定一个非形而上学的基础。无论结论如何，只要是在非形而上学的基础上得出的，就可以在共同的基础上进行争论，取得一致意见，而形而上学的理论对立，则是无法争论的，不属于科学的范围。

5.4 美学的情绪基础

“美是什么？”是一个众说纷纭的问题。就结论来说，我们基本上同意某些哲学家的观点：“美就是善”。在本书的 2.3 节我们把真理归结为善理，在本节中我们将论述“美”的概念最终也可以归结为“善”的概念，“善”是“美”的深层基础。13 世纪的神学家托马斯·阿奎那 (Thomas Aquinas) 认为，美的事物是一种在人们看见它时能给人以快乐的事物 (*The beautiful is that which pleases us upon being seen*)。按照这个定义，美就意味着感官的愉

悦，而从进化和认知的角度看，人之所以对某些事物感到愉悦，是由于这些事物对人的生存具有正的价值，或者说对人来说是善的事物，是值得追求的事物。如果美的感觉功能不符合善的要求，那么这种美感应该在进化过程中被淘汰掉。

但是，如果把“美”的概念与“善”的概念等同起来，“美”作为一种观念就不再具有独立的意义，美学也就没有存在的必要了。为了区别美和善的概念，康德（Immanuel Kant）对美附加了一个特征，即美给人带来的快乐必须完全是“无利害关系的（*disinterested*）”快乐。如果接受康德这种关于美的定义，会出现一个困难的问题，即我们无法说明人类的审美能力为什么能够在进化过程中产生并被保存下来。因此我们需要对康德的定义作一点修正，即美给人带来的愉悦并不是完全无利害关系的，而是缺乏直接利害关系的，从长远和综合的角度看，仍然是有利害关系的。人的美感，对某一事物产生愉悦的心情，是对该事物在感情评价系统中的正评价引起的。这种正的评价并不是由事物本身产生的，而是由该事物所具有的特征产生的，一些事物本身并不具有多大的利害关系，但是它所具有的某些特征却是很多对我们有利的事物所具有的共同特征，那么这类特征从统计意义上带有正的评价，因而可以引起愉悦的感觉。

例如，通常我们对彩色饱和度比较高（色彩鲜艳）的和形状规则的物体容易产生美感，是由于在自然界中生物比非生物通常具有较高的彩色饱和度和更规则的形体，而对于人类生存来说生物的意义显然大于非生物的意义。因此这两种特征具有较高的价值，可以成为引起愉悦感觉的因素。而非生物的宝石，由于具有相同的特征，也能够引起美的感觉。作为宝石的各种因素都是含有正评价的特征，首先是色彩鲜艳，再就是硬度高，硬度高意味着不易被破坏，具有较高的稳定性，符合存在原则，开采出的原石还要经过加工，使其获得规则的形状。我们在猎豹奔跑动作或

优秀运动员的泳姿中可以发现一种美，是因为这些动作符合经济原则，用最少的能量获得最高的速度。虽然宝石不能吃，猎豹用什么姿势奔跑与我们并无利害关系，但是由于它们具有使我们产生正评价的特征，从而可以使我们获得美的感受。

预测原则是在学习中起重要作用的原则，而评价又是学习所必要的导引信息，因此可预测性是美感的一个重要来源，主要体现为秩序的美感。任何乐曲都是有节奏的，节奏使得乐曲具有可预测性和秩序的美，完全失去节奏的乐曲听起来不象乐曲而更象噪音，给人一种混乱的感觉而失去美的享受。在自然界之中我们很少见到直线，但是在人工建造或制造的物体中直线是主要的几何要素，这不仅由于制造方便和力学结构上的合理性，以及两点间直线最短而体现的经济原则，还由于直线是最容易由其一部分预测其他部分的一种空间曲线，因而能够由于符合预测原则而带来美感。规则的几何图形和图案的美感也是由其可预测的秩序性产生的。对称的美也是基于预测的，知道对称图形的一部分可以很容易推测与之对称的部分。当然对称还有更直接的利益，所有能快速移动的动物其外部形状都是左右对称的，否则在快速移动时难以保持直线运动，因此我们的交通工具无论是车辆、船舶、飞机都采用左右对称的形状。因此，美的感受并不是直接由某种利益引起的，但是引发美感的特征肯定是与某种利益密切相关的。

与预测原则有关的不仅仅是秩序性的美。由于预测原则在认知学习过程中的作用，人的认知系统为了通过学习增强其预测能力，把原来不可预测的现象变成可预测的，就有必要对暂时不可预测的现象格外关注，这种要求在审美感觉中的表现就是新颖性的美，也可以称之为时装性的美。由于认知机制中短桥原理的限制，我们往往难以接受过于怪异的东西，至少对于象白噪声这种永远新颖而且永远不可预测的东西，人们是不会感兴趣的。如果我们到电影院去，结果发现上映的电影是过去看过的，会由于缺

乏新颖性而感到失望,但是如果是在电影放映过程中突然停电,那么这种过度的新颖性可能使我们更加失望。因此能带来美感的新颖性,只局限于我们的认知机构所能够理解的范围之内。当我们对新颖的事物逐渐熟悉起来,新颖性也就会随之消失,对该事物的美感也就会下降,只保留由其他因素所引起的美感。实际上,只由新颖性是不能产生美感的,第一次见到的丑的事物我们会觉得更丑。因此,新颖的作用是强调和放大,由于新颖性的吸引注意力的作用使秩序美等其他因素造成的美感效果更为强烈。所谓“第一印象”或“一见钟情”就是基于这种新颖性的增强效果。

与文化的运行和进化机制有关的是接受性美学。接受性美学是指人类受文化基因控制的审美评价准则会受到文化环境的强烈影响。在人的审美准则中有源于生物基因的,也有源于文化基因的。个人对于文化环境中审美准则的接受倾向,与其他方面的文化接受能力一样,是有利于个人生存的。个人的审美准则及其利害评价准则与文化环境中多数人相互一致,会使文化环境对于个人来说更容易预测,也使得个人更容易被其他人所理解和容纳。基于这种一致性所具有的利益,使得人在评价准则方面具有求同倾向和接受能力,因而使得一些新的文化产品会发生“流行”的现象。流行的初期动力是新颖性的作用,后期则借助于接受性美学。流行歌曲,时装,乃至流行的理论,其流行机制大体都是如此。流行现象在文化进化中所起的作用是使少数人创造的文化基因(变异)传播于文化群体,接受广泛的选择,属于文化进化的软结构;未经受住选择考验的文化内容将被新的流行所取代,而经受住选择考验的部分则淀积下来,形成对文化硬结构的建构。

关于美学还有一个重要的问题,美是主观的还是客观的。从美是人的一种感受来看,美是主观的,没有人的观察,就没有美丑可言。但是如果认为美是纯主观的而毫无客观性,那么美学也就成了个人宣称的记录,不可能成为学问。美的客观性就在于,人

们可以在关于美的问题上取得某种程度的一致意见。与“善”相比，“美”的客观性要差一些。因为存在原则和经济原则可以不依存于人的认知结构，而美的感觉产生于对某些特征的正向评价，特征抽取又是人类认知机制的一部分，所以美比善的主观性更强。但是美仍然具有一定的客观性。美的客观性的根据在于，人类基因中硬基因的一致性使得人的审美能力具有相同的生物学基础，在人的审美准则中基于生物学的准则是大体一致的，这种一致性保证了在人的审美准则中有一部分是全人类一致的。此外文化中的求同倾向使得在一个文化群体之中多数人会具有类似的审美准则。生物与文化都处于进化过程之中，因此无论在生物基因和文化基因中都会存在着变异成分，体现多样化的价值。因此人的审美准则也和其他评价准则一样，存在着个人差异，这就是美在一致性意义上的主观性。

5.5 社会的通用评价尺度 ——论金钱

“金钱不是万能的，但是没有金钱是万万不能的。”自古以来关于金钱的议论众说纷纭，有时候金钱被看成是压倒一切的力量，有时候金钱又成了挡路的东西“阿堵物”。在金钱的作用越来越强的今天。金钱是什么？金钱到底起什么作用？这样的话题在文学作品中也是常常出现的。金钱的形式是千变万化的，贝壳、铜币、贵金属、纸币、磁卡、直到银行计算机中的数字，金钱经历了一个逐步抽象化的进化过程。最早的金钱只是物物交换过程中的一种物品，用于金钱的物品需要满足一些条件，即便于保存、便于携带（流通）、易于精确计量、易于精确地表现计量（可分性）以及安全性等等。按照这些条件，就很容易理解金钱的进化过程，贝壳显然是不便于精确计量的，金银便于携带和保存，但可分性差，

不便于小额使用，所以通常只用于大额计量，和小额的铜币共同流通，纸币可以比较好地满足各项条件，是使用范围比较广的货币，计算机的发明使得货币又进入了一个新的时代，计算机中的数字可以使与货币有关的一系列处理自动化。由货币的进化过程可以看出，金钱的意义不在于它的物质性，而在于它的信息性，或者说，金钱不是物质，而是一种信息，因此，货币的流通并不限于把金块或纸币从一个地方搬到另一个地方，而可以象其他信息一样，通过通讯系统的传送来实现。

金钱是一种信息，这种信息的主要作用是作为社会通用的评价尺度，使社会范围的价值评价得以量化。对于一个脱离社会而生存的单独的个人来说，金钱是没有用处的，他完全可以按照头脑中的生物价值体系来决定该干什么和不该干什么。金钱起源于社会中的交换行为。交换的基本原则不是等价交换，而是不等价交换，或者称之为增值交换。以一个简单的模型为例，甲乙双方持有不同物品，甲持有物品 A ，乙持有物品 B ，甲对物品的评价是 $B > A$ ，而乙对物品的评价是 $A > B$ ，此时甲用持有的 A 与乙持有的 B 进行交换，通过交换后对甲和乙来说价值都增加了。这就是交换的增值原则。增值的背景是社会分工与合作造成的利益，这种利益通过交换而实现。如果交换不产生增值或反而减值的话，交换就会失去意义。交换行为本身也是有成本的，因此在交换中必须增值并且增加的价值应大于交换成本。例如花上一天时间走几里路去赶集，就构成交换成本。当交换不是直接的以物换物，而是通过货币这一中介物间接进行时，增值交换原则依然成立，即只有当人们对商品的评价高于对商品价格相等的货币的评价时，购买行为才会发生，或者说，我们只买值得买的东西。

在一个人脱离社会独自生存时，他作出决策只需要依据自己的认知系统中对各种不同需求和支付劳动的评价比较。由于需求是多种多样的，支付代价也是多种多样的，决策比较必需对多样

的价值进行统一度量，否则无法对不同质的量进行对比。在认知系统中，这种统一的量可以是众多相互抑制的神经模块的兴奋程度。食欲、性欲、危险、劳累等诸多因素最终都化为神经元兴奋强度这个统一尺度来进行对比和决策。有了这个统一尺度，头脑中的综合决策才能进行，使得头脑这一极复杂的系统实现决策的最佳化。金钱对于社会的作用，也是在复杂系统中作为统一的评价尺度，以利于社会总体决策的最佳化。当个人生活在社会中时，可以享受分工合作的利益，通过满足他人的需求来达到满足自身需求的目的，但他人的需求或社会结构所包含的他人利益是非常复杂的，个人的认知结构中所包含的价值体系通常不具备判断他人需求或社会结构需求的能力，因此金钱就成为个人判断社会需求的重要评价信息。反之，社会对某种需求的评价，最终是依赖于组成社会的个人对某种需求的评价。因此，商品的价值主要取决于消费者的评价，而不是生产成本。某种商品生产能够成立，必须是其生产成本低于由消费者评价所得的价值。在自由加入的市场上，市场机制平衡的结果可以使商品的价值接近于生产该商品的社会平均劳动时间所代表的价值，但这是平衡的结果，而不是价值产生的原因。如果用生产成本来规定价值，就无法解释提高社会劳动生产率对社会优化的贡献。社会的评价归根结底取决于社会中每个个人的生物价值体系。

有些道德家看不惯市场上的“砍价”现象，他们认为，还价的可能性，意味着卖方最初的讨价高于商品的实际价值，是一个谎价，因此属于欺骗行为。如果我们认为商品价值由生产成本决定，那么道德家的看法是合乎逻辑的，此时卖方的正当行为应该是按商品的进货价格加上商业成本再加上适当的利润，报出一个合理的价格。问题在于“适当的利润”是卖方所无法估计的，也是卖方单方面不可能决定的。如果用增值交换的眼光来看待交换和交易，就很容易理解讨价还价的意义。例如对于一件衣服，卖

方认为卖 100 元可以保本，买方认为最多值 200 元，否则不值得买，那么就构成了增值交换，交易是可以成立的，实际成交价格 在 100 元和 200 元之间的任何一个值，对双方都是有利的。讨价还价所解决的问题仅仅是怎样分配由交换产生的利益，不存在谁欺骗谁的问题。上述的例子如果以 140 元成交，那么卖方获得的利润是 40 元，买方的交换增值是 60 元。由于买卖双方都力图获得更大的交换增值，通常也都不知道对方的评价，买方可以尽量造成低评价的假象，卖方也可以造成高成本的假象，作为讨价还价的战略。博弈论关于讨价还价的研究表明，力图获得最大交换增值的战略有时会使本来可以成交的交易失败，使双方都失去了交换增值所产生的利益。因此合理的价格意味着交换增值的最大化与交易失败风险的平衡。讨价还价本身要付出时间和精力，也构成一种成本，在讨价还价带来的利益不大于成本时，讨价还价是没有意义的。在没有讨价还价的交易之中，买方的评价仍起主导作用，表现在买或不买的决策过程之中，没人买的商品最后只得削价处理，这也是隐含的讨价还价机制。因此，客观的价格最终还是多人的主观评价在市场机制中作用的结果。

个人在市场经济中通过正当竞争获得巨额财产，表明他能够正确地掌握社会需求和普遍评价，从而获得了社会对其贡献的承认，这就是“大款”和“大亨”的社会意义。一个人善于使所控制的金钱增值，意味着他具备判断社会价值的决策能力，而社会中的金钱通过增值向这些人的控制领域流动，又会使更多的社会财富得到更有效的利用。这就是整个社会行为优化的一种机制。一个人拥有巨大的财富，并不意味着他可以消耗掉这些财富，他需要执行社会财富增值的使命，否则他将不再拥有。即使他把一些财富用于挥霍，客观上仍是为他人创造赚钱的机会，吃一顿 1000 元一人的宴席并不能比吃 10 元的便饭多吃价值 100 倍的东西，只是为饭馆创造了更多的利润。一个人所能消耗的社会财富是极为

有限的。可以说，合法致富的大款是社会优化的功臣，或者说大款是金钱增值本性的驯服工具。

以上的说法只有在正当竞争的前题下才能成立。如果在追求利益时对于手段不加任何限制，那么在追求利益的过程中会发明出一些破坏市场机制的手段，使市场机制失效，此时追求利益不再具有使社会整体行为优化的作用。这就象足球赛是通过把球射入对方球门的目标而达到提高球技的目的，但如果对方法不作任何限制，足球可能转化为武术或枪战，足球作为一种运动也就灭亡了，射门也失去了意义。不正当竞争的方法是很多的。有些方法的不正当性比较明显，例如制售伪劣商品，其直接作用是使消费者对商品价值产生错误判断，远期结果是降低市场作用，增加市场的整体运行成本。以权谋私和收回扣的行为，会导致社会腐败、使社会结构偏离或丧失其设计功能，导致社会劣化。垄断是竞争机制的自然产物，其结果又是取消竞争和市场机制，这是竞争自身的悖论性质，因此需要法律的外力防止垄断的产生，而倾销则是制造垄断的方法，也属于不正当竞争之列。虚假广告也是诱使消费者对商品作错误评价，其性质与伪劣商品相同。

有些不正当竞争的方法比较隐蔽，不通过分析不易看破其实质，这种不正当竞争的典型例子是重奖销售和老鼠会。重奖销售的不正当性在于使商品销售的性质发生了变化，使消费者不是根据对商品的评价及价格来选择商品，而是在购买随商品而来的中奖机会，实质上是把销售商品变成了销售彩票。彩票是利用赌博心理赚钱的方法。由于赌博心理的作用很大，各国都对赌博有严格的限制，发行彩票通常只作为社会公益事业集资的手段，而不允许作为私人或集团谋利的手段，因为这种私人谋利的方法对社会的优化毫无作用。从博弈论和统计学的角度看，买彩票是不上算的，因为彩票发行者的稳赚在零和博弈中意味着买彩票的人一定吃亏。但是人的价值体系往往看不清这种吃亏，这是由于人对

于不同的尺度敏感程度不同，人们会感到一百万元比一万元多得多，但不容易感觉出百万分之一的机会比万分之一的机会少得多，在一般人看来，二者都是“很小的”机会。因此中奖金额越高的彩票越吸引人，而中奖概率的下降却不易引起人们注意，这是由于人们在日常生活中对于微小数字是不大在意的。此外，生物的进化本身是一场大赌博，在突然变异之中产生有利变异的机会只有一亿分之一，但是基因还是不断产生变异，否则就完全没有进化的可能性，进化过程是一个极不经济的过程。在社会中有成千上万作明星梦的人，而只有极少数人真能当上明星，但是如果没有大量的破梦者，也不可能产生好的明星，这是赌博心理的正面意义。不过基因的变异不可太多，以至于导致生殖能力下降，作明星梦也不应断了其他的生路，用不关痛痒的金额去买彩票是正常的行为，而用全部财产去买彩票则是荒唐的，把全部彩票都买下来的人肯定中奖，但他也肯定亏本。

老鼠会是一种更为隐蔽的不正常竞争方式，往往打着“直销”的旗号。“老鼠会”这个词来源于一个数学题：一对老鼠每月生8只老鼠，出生后两个月的小老鼠又开始产仔，一年后一共有多少只老鼠。老鼠会的运行机制与老鼠问题相似，一个人花高价买取一定数量的商品，从而获得推销员的资格，向其他两个以上的人高价推销相同商品，获得一定的提成来补偿自己购买高价商品的损失，甚至获得一定的利润，而下一代的购买者又成为推销员，一代一代推下去。这种“直销”方式的欺骗性在于，从表面上看，每一个参加老鼠会的人都不会亏本，甚至可以赚钱，但是这种情况只有在需求无限的条件下才能成立，而任何市场都不会是无限大的，老鼠也不可能无限增加，后出生的老鼠命里注定要被饿死，而市场的饱和又是非常快的，即使是10亿人的市场，一个人传两个人的老鼠会在不到30代就可以撑破这个市场，这就是指数爆炸的威力。而在市场饱和时，按照零和博弈的原则和指数

计算都可以得出，结局必然是挨坑的人比赚钱的人多。老鼠会方式的不易识破之处在于，挨坑的人虽然吃亏，却以为是自己推销的能力或努力不够，看不出这个系统的设计就是一个保证多数人吃亏的圈套，即使推销成功，也只是增加上当人数而已。老鼠会发展越成功，上当者的总数越大。因此在一些国家有专门的法律条款禁止老鼠会，其性质和诈骗是相同的。

以上关于金钱的评价尺度作用的讨论是以市场经济为背景的，然而经济的形式是多种多样的，计划经济也是一种经济形式，在特定的条件下还是一种有效的经济形式。中国正处于从计划经济向市场经济的转轨过程，讨论计划经济与市场经济的区别是很有意义的。计划经济在一些国家的历史上曾经取得过辉煌的战绩，而在一定时期之后又都暴露了其局限性，那么计划经济在什么条件下是可以成立的呢？从根本上说，计划经济是建立在充分预测的基础之上的，而市场经济则是一种进化机制，从原则上是不可预测的，它实现的是自寻最佳点的过程。另一方面，市场经济是需求导向的，而计划经济本身不包含需求导向的机制。从以上区别可以看出，计划经济在后发工业化国家的建立工业基础阶段是优于市场经济的。工业化的先驱国家面临全面的不可预测性，只能采用市场经济的方法，而后发国家由于有经验可借鉴，有可能对发展前景作出一定的预测，靠设计而不是进化机制前进，进化的探索过程是很不经济的。为了加快工业化过程，需要采用低消费、高积累的方式获得资金，在这种情况下，消费需求的导向作用反而是不利的，会导致有限资金的分散。此时，社会建立工业化基础的长远整体利益要求个人牺牲眼前的消费利益，而生物和社会的自然进化又都是看重局部和眼前利益的，因此抑制消费欲望和取消进化机制是有利的，艰苦朴素成为这一时代的道德要求具有必然性。受到充分抑制的基本生存需求也是容易预测的。当社会工业基础建立，生产发展到一定水平时，社会需求超出基本

生存需求而变得复杂起来，可预测性下降，计划经济缺乏进化机制的弊病开始显露出来，经济成长也开始下降。社会需要用市场机制自行优化，需要用消费作为优化的导向。在从计划经济向市场经济转化的过程中，消费逐渐成为一种美德，过度抑制消费，会使经济发展失去方向。当然，艰苦朴素仍然是重要的，从社会整体看，适度的高储蓄率有利于扩大再生产，从个人来看，习惯于低消费水准生活可以拓宽自由选择的范围，有利于长远打算。

把金钱作为社会优化所需要的通用尺度，可能是很多人难以接受的。一般社会常识认为，有很多价值是金钱所无法度量的。但是在现实之中，人们经常会面临金钱与道德之间的平衡决策问题，此时道德并不具备形而上学的无限大价值。当我们对于道德的价值有了更准确的估计时，决策也会更加合理。见利忘义的错误在于过小的估计了道德的价值，其结果是得小利而失大利。反之，刘邦如果把“孝”看得过重，不肯为争天下而用父亲的生命作赌注，也就不可能成为汉高祖。在经济发展与环境保护之间也存在类似的矛盾，但是如果通过科学研究能够对环境破坏的损失作出定量的分析，对于这一问题也可以通过计算作出合理的决策，例如决定到底应该先污染后治理，还是应该适当牺牲发展速度而从一开始就保护环境。金钱在这类问题上仍然可以成为通用尺度，以利于进行理性的决策，使决策从直觉上升到科学和定量的水平。



参 考 文 献

- [1] Marvin Minsky, THE SOCIETY OF MIND, 1985.
- [2] Douglas R. Hofstadter, METAMAGICAL THEMAS, Basic Books Inc. 1985.
- [3] 道格拉斯·R·霍夫施塔特、丹尼尔·C·丹尼特,《心我论》,上海译文出版社,1989。
- [4] Howard Gardner, THE MIND'S NEW SCIENCE, A History of the Cognitive Revolution, 1985.
- [5] Donald A. Norman, PERSPECTIVES ON COGNITIVE SCIENCE, Ablex Publishing Co., 1981.
- [6] Karl R. Popper and John C. Eccles, THE SELF AND ITS BRAIN, Springer-Verlag, 1977.
- [7] [英] 布赖恩·马吉 (Bryan Magee) 著, 南砚译,《“开放社会之父”——波普尔》, 湖南人民出版社, 1988。
- [8] John Watkins, SCIENCE AND SCEPICISM Princeton University Press, 1984.
《科学与怀疑论》, 邱仁宗、范瑞平译, 上海译文出版社; 1991。
- [9] Douglas R. Hofstadter, G ÖDEL, ESCHER, BACH. Basic Books Inc., 1979.
《GEB——一条永恒的金带》, 乐秀成编译, 四川人民出版社, 走向未来丛书, 1984。
- [10] John Maynard Smith, EVOLUTION AND THE THEORY OF GAMES, Cambridge University Press, 1989.
- [11] 钱学森主编,《关于思维科学》, 上海人民出版社, 1986。

[General Information]

书名=认知科学与广义进化论

作者=赵南元

页数=364

SS号=10199586

出版日期=1994年03月第1版

封面

书名

版权

前言

目录

第一章 绪论

- 1.1 三个许诺和一则寓言
- 1.2 人工智能中的问题及其困难之处
- 1.3 认知科学的方向

第二章 工程认识论

- 2.1 东西方文化比较和深层精神结构
- 2.2 科学与形而上学
- 2.3 从真理到善理
- 2.4 认识论的逆镜模型

第三章 广义进化论

- 3.1 自我表述系统
- 3.2 进化的地形图模型
- 3.3 进化论的博弈论方法
- 3.4 广义进化论的软硬结构模型
- 3.5 生物的进化与进步
- 3.6 有性生殖的意义
- 3.7 广义进化范式

第四章 认知与思维

- 4.1 认知的阶层性
- 4.2 神经网络
- 4.3 神经网络群体
- 4.4 从动物到人的遗传与学习
- 4.5 智能模型与形象思维
- 4.6 感情与评价
- 4.7 意识与自我
- 4.8 创造与灵感
- 4.9 模式识别
- 4.10 人工认知主体

第五章 广义进化系统

- 5.1 复杂巨系统
- 5.2 全局观点与分离主义
- 5.3 伦理学的非形而上学基础
- 5.4 美学的情绪基础
- 5.5 社会的通用评价尺度 论金钱

参考文献