



中青年学者文库

程 炼 © 著

思想与论证



北京大学出版社
PEKING UNIVERSITY PRESS

思想与论证

Thought and Argument



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

思想与论证/程炼著. —北京:北京大学出版社, 2005. 10

(未名·中青年学者文库)

ISBN 7-301-09285-7

I. 思… II. 程… III. 分析哲学—研究 IV. B089

中国版本图书馆 CIP 数据核字(2005)第 067364 号

书 名: 思想与论证

著作责任者: 程 炼 著

责任编辑: 赵永清

版式设计: 王炜烨

标准书号: ISBN 7-301-09285-7/G·1535

出版发行: 北京大学出版社

地 址: 北京市海淀区成府路 205 号 100871

网 址: <http://cbs.pku.edu.cn>

电 话: 邮购部 62752015 发行部 62750672 编辑部 62750673

电子信箱: xuyh@pup.pku.edu.cn

排 版 者: 北京高新特打字服务社 82350640

印 刷 者: 三河新世纪印务有限公司

经 销 者: 新华书店

650 毫米×980 毫米 16 开本 15.75 印张 230 千字

2005 年 10 月第 1 版 2005 年 10 月第 1 次印刷

定 价: 32.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,翻版必究

说明和致谢

本书的15章中,有9章是我过去10余年里发表的单篇论文和较长的书评。在准备本书的过程中,我对其中的大部分做了修改,许多改动是剧烈的。我的修改试图遵循三个原则:(1)剔除我现在看来是明显错误的内容;(2)补充我现在认为必要的新内容和论证;(3)维护全书各章观点之间的一致性和连贯性。其余6章都是新作尚未发表的。

下面列出曾发表的旧作的出处,并向原出版机构致谢:

第1章“人工智能基础评论”,原始版本以同标题发表于胡文耕、吴国盛主编的《科学前沿与哲学》,北京:中共中央党校出版社,1993年。

第2章“第一人称哲学的局限”,原始版本以同标题发表于赵汀阳主编的《论证》1999年秋季号,辽海出版社,1999年。

第3章“论人是机器”,最早的版本以“人机类比的含义”为题发表于《科学时报·读书周刊》2000年5月22日;后以“人是机器又何妨?”为题发表于《哲学门》第二卷第一册,湖北教育出版社,2001年。

第5章“私人语言论证”,原始版本以同标题发表于赵汀阳主编的《论证》1999年秋季号,辽海出版社,1999年。

第6章“先验论证”,原始版本以同标题发表于《哲学

研究》1998年第10期。

第7章“刘易斯与怀疑论”，发表于《云南大学学报·社会科学版》第3卷(15)，2004年第6期。

第8章“尼采的挑战”，发表于《科学文化评论》第1卷第6期，2004年。

第9章“蒯因的本体论”，发表于《外国哲学》第16辑，北京：商务印书馆，2004年。

第11章“论自由”，发表于《云南大学学报·社会科学版》，第2卷(8)，2003年第5期。

许多同事和朋友以各种方式为本书的完成提供了帮助和支持。孙永平阅读了第2、3、4、5、6章的手稿，订正了许多错误，徐向东阅读了第7、9、12章的手稿，提供了许多重要的修改意见。谨向他们致以谢意。邢滔滔阅读了第7、8、9、12、13、14、15章，提供了许多书面评论，他多年来一直是最有耐心的读者和批评者，我现在还在困惑，他如何能够不厌其烦地阅读我一遍遍不停修改的稿子。也是他，激发了我对宗教哲学的兴趣，没有他，本书将只有前四个部分。

第13、14两章脱胎于分别提交给两届中美哲学论坛(The Sino-American Philosophical Forum:2002年清华大学,2004年北京大学)的会议论文。我感谢出席会议的美国哲学家们的评论，他们中有：凯利·詹姆斯·克拉克(Kelly James Clark)、迈克尔·穆瑞(Michael Murray)、里克·奥特(Ric Otte)、艾尔·普兰廷伽(Al Plantinga)和梅尔·斯图沃特(Mel Stewart)。

北京大学出版社杨书澜女士为本书的完成和出版付出了艰辛的工作，在此我向她致以深深的谢意。

靳希平教授和张祥龙教授，我的同事、朋友和榜样，多年来一直在各方面给予我支持和鼓励(当然也不乏批评)，我无以为报，只好把这本似乎属于分析哲学的微薄的小书，献给这两位出色的现象学家。

程 炼

北京，2005年2月

目 录

导言 从苏二条看哲学	(1)
第一部分 心灵	(5)
1. 人工智能基础评论	(5)
2. 第一人称哲学的局限	(29)
3. 论人是机器	(43)
第二部分 语言	(58)
4. 悖论与实践	
——维特根斯坦论遵守规则	(58)
5. 私人语言论证	(73)
第三部分 知识	(91)
6. 先验论证	(91)
7. 刘易斯与怀疑论	(107)
8. 尼采的挑战	(118)
第四部分 世界	(128)
9. 蒯因的本体论	(128)

10. 生与死	(149)
11. 论自由	(168)
12. 休谟的因果性说明	(181)
第五部分 上帝	(203)
13. 道德神令论的难言之隐	(203)
14. 普兰廷伽与自然主义	(214)
15. 成乎败乎改革宗?	(229)
参考文献	(242)

导 言

从苏二条看哲学

人类是通过官能,同世界打交道的。有些人用色彩和线条表达自己和世界以及它们的关系,如果人类天生是盲的,这些做不成;有些人用音符、旋律和节奏做这些事情,如果人类天生是聋的,这些也做不成。哲学家做这些事时用的是认知上的官能(cognitive faculties),没有这个官能,他们也不能成事。哲学起源于对世间万物(包括人类自身)迷惑和好奇,人类用自己的认知官能来破解迷惑,满足好奇,因此,这是一种智力上的活动,不同于吃喝拉撒睡这些生理的活动。哲学家做事情的工具是概念,一种抽象的东西,他们用概念来表达困惑、对它们进行分门别类的研究,并构造各种理论来解决困惑。有趣的是,一旦他们认为某些理论已经充分好地解决了问题,它们就变成了科学。今天,科学的地盘越来越大,有些人认为,科学将最终给人类提供关于一切事情的说法。人们会问,那哲学干啥?

其实,大可不必担心哲学会像蝶蛹一样,一旦化蝶成蛹就没用了。人类的迷惑和好奇是没有止境的,科学大大成功,但哲学的困惑依然存在,因为在很大程度上,哲学的提问有惊人的超越性和一般性。比如说,当一个数学难题被解决,数学家认为他们获得新知识时,哲学家会问:知识到底是什么?考古学家断言一件物品是某个

时期的,哲学家会问:时间到底是什么?当你说宇宙是大爆炸产生的,哲学家会问:大爆炸有没有原因?当你说撒谎是不对的时候,哲学家会问:什么使得一个行为有对错?当你说“单身汉”的意思是“未结婚的男人”时,哲学家问:什么使得一个词有意义?当你说某人有一颗“美丽的心灵”时,哲学家问:心灵是个什么东西?当你说吸烟引起肺癌时,哲学家问:什么是因果性?任何事情都有原因吗?某些人认为有超自然的东西,如菩萨、上帝等,哲学家问:这些想法是合理的吗?

所有这些问题,都是典型的哲学问题、传统的哲学问题。它们中的绝大部分打从哲学诞生时起就一直没有消失过。一代一代的哲学家不停地寻求对它们的更好的回答。哲学问题的顽固性本身就是一个很严肃的哲学问题。有些人从这种顽固性中得出哲学问题不可解的结论;有些人干脆认为哲学问题是虚假的、无意义的;还有一种被称为“后现代主义”的思潮充斥着今天的世界。后现代思潮有形形色色的变体,有的人说哲学只是西方文化的产物;有的说哲学是人类思想的一种病态;有人说哲学死亡了、被超越了、终结了;如此等等,不一而足。这里显然不是详细评论这些说法的地方。不过,我愿意指出的是,既然它们都不是说着玩的,既然它们构成了一套说法,那么它们就必须满足一项要求,那就是,理智上的可接受性。这个要求是否被满足,不能光看这些说法有多流行、有多少附和者,而要看它们在理智上的质量是否合格。后现代主义者说传统哲学消亡了、终结了,除非他们能提供在力度、深度和真度都不差的替代品来,否则他们就有点哗众取宠、耸人听闻。

理智上的可接受性,是一个规范的要求。这项要求一方面是外部的,也许是我们意识不到、观察不到的,那就是,我们的认知官能必须运转良好,我们是在恰当地、充分地运用自身拥有的能力。另一方面,这项要求已经体现在一个由苏格拉底集成、继之以亚里士多德、阿奎那、笛卡尔、莱布尼兹、休谟、康德、弗雷格、罗素等等(恕我实在不能列全)的伟大传统中,这是我们能够意识得到、观察得到的。苏格拉底身上显示出哲学探索的两项指标(我把它们简称为“苏二条”):第一条,我们要用清晰的概念将思想清晰地表达出来,我们要用字面的意思(literal meaning)而不是修辞的方式说

话;第二条,我们的思想要经得起推敲和论证,而不是愿望式的、跳跃式的、故弄玄虚的。因此,从苏二条看哲学,哲学乃是概念之学、论证之学。在这个意义上,我认同的是一个唯智主义(intellectualism)传统,我的工作乐于接受这个传统的约束。反对苏二条,就是反智主义。反智主义者在这个世界上比比皆是,想必他们不屑于争抢哲学这块招牌。

这并不是说,只要遵循了这两个指标,就足以产生确定的哲学结论。哲学问题的顽固性在很大程度上是由哲学结论的不确定性造成的。这不难解释。哲学既不同于有形式的方法可以遵守的数学,也不同于有实验数据来进行验证的自然科学。从哲学史可以看出,绝大多数哲学学说缺乏广泛的可接受性,原因就在于,它们或者受制于概念的不清晰,或者受制于论证的不严密,或者受制于两者。苏二条虽然看起来简单,实践起来是难上难。

美国哲学家托马斯·内格尔(Thomas Nagel)有一句耐人寻味的话,“哲学是智识(intellect)的童年,一个试图跳过它的文化永远无法长成。”^①这个童年要成长,需要什么营养?曰科学,曰逻辑。哲学若是概念之学、论证之学,何处寻找概念、怎么增强论证?罗素曾说,哲学介于科学和神学之间。他强调了科学与宗教的对立,并力主哲学应当与科学结盟。在我们这个宗教气氛不浓的国度里,我宁愿说哲学在科学和常识之间。与罗素一样,我也坚信哲学应当与科学结盟。其实在我看来,哲学探索和科学探索本来就是紧密相连的。但我也听到另一些抱怨,说如果这样哲学(甚至包括科学)太去魅了,远离了我们的生活,忽视了生动的直观,遮蔽了大地、神庙、存在的意义……抱怨者们把“生活形式”、“生活世界”、“现存在”等当作更基本的哲学概念(或许他们不愿意把这些称为“概念”),要还魅,让世界充满魅力。这些很“魅”的说法,对许多人特别对哲学学生和人文学者来说,的确很“魅”,至少比做逻辑题、思考囚徒悖论、给图灵一丘奇论题找反例有“魅”的多。在我看来,苏二条是真正地照“魅”镜(恕我在这里用了修辞)。

^① 内格尔:《无源之见》(*The View From Nowhere*, New York & Oxford: Oxford University Press, 1986),第12页。

我说哲学的营养来自科学和逻辑,并不是说哲学完全是被动地接受科学结论,当然不是,也不应该是。科学和逻辑不是铁板一块,一成不变的,它们也在向更正确的形态发展。哲学的成长与科学的发展从来就是同步的。柏拉图、亚里士多德、笛卡儿、莱布尼兹、弗雷格、哥德尔不仅是整个人类历史上最出色的哲学家,而且还是他们自己时代最优秀的科学家。在我看来,我们这个时代的智识方面最大的遗憾之一是,名声最大的哲学家在科学上几乎一无所成。

本书的15章,每章都有相当大的独立性,读者可以按自己的兴趣去选择阅读。我按照主题将它们大致分成5个部分,5个部分分别论及心灵的特点、语词的意义、知识的地位、世界的结构以及宗教信念的性质。每篇文章前面我都加了一段简短的提要。需要注意的是,这个划分是松散的、完全出于方便考虑的。在我看来,哲学问题从来就不是一个一个孤立地存在的,对任何一个问题的处理都应该反映出一种体系性的思考。我期待读者用苏二条来照我的这件作品。

第一部分 心 灵

1. 人工智能基础评论

内容提要：人工智能的基础问题涉及人工智能的可能性和局限性问题。图灵的智能理论和机器思维的观点在本章中得到阐发。本章的主要内容是回应两种在哲学文献中反复出现的论证，一个应用哥德尔定理，一个运用现象学学说，两者都声称计算机能力与人类智能之间在原则上存在着不可填补的沟壑。本文试图证明，这两个论证都没有成功地证明这个沟壑的存在。

什么是人工智能？

人工智能最初是作为计算机技术（尤其是软件技术）的一个分支而出现的，经过几十年的理论与实践活动，取得了巨大的成就。在其发展过程中，哲学、心理学、语言学、神经科学等领域的大量思想和方法不断地渗入进来，以至于今天询问“究竟什么是人工智能”时，即使让有权威的专家们来回答，答案也是不尽相同的。

根据某种普通的定义,人工智能是使计算机做聪明的事情的艺术。这个定义虽然看起来抓住了人工智能的精神,但它不够具体,还有一些别的看法:

(1) 作为软件工程的一个分支,人工智能仅仅是一些编制程序的技巧,它们使得计算机能够诊断疾病、理解自然语言等等。

(2) 作为一门计算机科学理论,人工智能是关于编程的一种独特构想。

(3) 作为哲学的分支,人工智能是一种实验认识论:知识是什么?在计算机或人的心灵中知识如何被表达?

(4) 作为一门心灵(mind)科学,人工智能体现了这样一个思想:心灵基本上是一种处理信息的机制。在最深的层次上,人工智能试图对这个伟大的未决问题做出贡献:心灵如何从非心灵中产生出来?

站在不同的角度看人工智能,它就会呈现出不同的图像。在上面的前两个定义中,人工智能被当作一门技术,而后两个定义将人工智能看作一种思想。但在种种差异背后隐藏着一种原始的一致性,即人工智能作为一种科学探究,它正在或试图达到什么目的。一般说来,贯穿人工智能的全部历史,它的目标在两个方面:

(1) 建造强有力的装置以尽可能多地(或全部地)完成通常认为只有人类才能完成的工作;

(2) 提出精致的、基于实证的新理论以解释人类的精神生活。围绕这两个目标所进行的工作正好分别构成了人工智能的实践活动和理论活动。

人工智能诞生的初期,在这两个方面都出现了令人瞩目的成果。许多研究者因此提出了更加雄心勃勃的设想,他们试图将人工智能看成是一门一般的智能科学,即看成是认知科学的核心。它的目标变成:提供能够解释(或许能够使我们复制)人类的全部心理现象的系统的理论。因此,在人工智能新定义中,“计算机”反而消失了。人们津津乐道的是一些抽象概念,如表达(representation)、计算(computation)、心理图像(mental image)等。

大多数研究者都承认,人工智能作为一门学科已经成熟,但是,那些乐观主义者的雄心勃勃的纲领能否实现,却不可能仅靠设

计计算机系统、编写程序来解决。有一类问题深深地与我们的哲学探究相联系着,它们肯定伴随着人工智能的全部历史困惑着我们,这就是人工智能的基础问题。简单地说,“人工智能的可能性与局限性”是这些基础问题的核心。

英国天才数学家阿兰·图灵(Alan M. Turing)于1950年发表的著名长文《计算机器与智能》,可以看作是人工智能基础研究的滥觞。^①在这篇文章中,图灵明确地提出了这样的问题:“计算机能思维吗?”为了回答这个问题,图灵设计了一种“模仿游戏”。设想有三个人,一个男人甲、一个女人乙和一个提问者丙。丙可以是任何一种性别,与甲和乙隔离开来。这个游戏的目的是让丙通过交谈确定甲和乙哪一个是男人,哪一个是女人。为此,丙轮流向甲和乙提出问题,甲和乙依次给予回答(问答可借助打字机进行)。如果在一段时间内,丙无法分辨甲和乙的性别,那么就在游戏中失败了;反之,则获胜。图灵设想,现在如果用一台计算机代替甲和乙中的一个,游戏就变成让丙来识别两个回答者中哪一个是人,哪一个机器。如果提问者无法识别,我们能说这台机器有智能吗?图灵的回答都是肯定的。

图灵的论点后来引起了广泛的争议,这里做一些澄清的工作是有必要的。我们把用图灵游戏来测定智能时所涉及的问题分为两个方面,一个是技术方面,另一个是原则方面。从技术方面看,图灵的原始论文在许多细节上是不清晰的。首先,模仿游戏需要进行多长的时间才算分出胜负,三、五分钟还是数日?如果时间太短,提问者从回答中得不出足够的信息,太长,机器可能死机,人可能累趴下。其次,交谈的内容是否有限定?第三,智力多少是个程度上的事情,某些人智力超群,另一些人愚不可及,更多的人处于中间地带。一台机器可能骗过一个智力平平的提问者,但在一个专家面前却过不了几招。第四,提问者的主观因素显然能影响到游戏的结局。我们是随意指定提问者,还是需要做一定的选拔?所有这些问题都能引发人们思考图灵游戏是否是一个切实可行的

^① 图灵:《计算机器与智能》(《Computing Machinery and Intelligence》*Mind*, vol. 59, No. 236(1950)),第433~460页。

方案。我无意在这里进一步讨论这类问题,因为解决它们依赖于更多经验探索。但是,我想指出的是,即使人们能够成功地就这些问题达成共识,还有另一类更理论性的问题没有触及,而这后一类问题是我更为关注的。

许多人以及大量哲学文献都注意到,图灵提出的智能观念是行为主义的。哲学上的行为主义主张,在内部的心理状态和外部行为之间存在着概念上的联系,也就是说,我们在谈论心理状态时,只不过是谈论行为或行为的倾向。但是,图灵的设想并非简单地建立在行为主义预设上。图灵对一般意义上的行为并无太大兴趣,他并未主张,如果一台机器在行为模式上与真人无法区分,那么这台机器就像真人一样具有心理状态,或者更具体地讲,具有智能。在图灵的模仿游戏中,机器只需展示一种特殊的行为——言语行为(verbal behavior),而在许多理论家看来,单有在语言方面与人类无法区分的表现,既不是一个东西具有智能的必要条件,也不是充分条件。实际上,图灵并没有主张,不能通过图灵试验的系统就是没有智能的。图灵的意思是,如果一个系统在一段合理长的时间内可以像一个人一样进行日常交谈,这就足以让我们断定它是有智能的,也就是说,在模仿游戏中获胜是一个系统有智能的充分条件。根据我们的日常看法,智能可以用许许多多方式体现出来,演算习题、下棋、创作文艺作品等都可以看作是智能的展现。为什么图灵对言语行为情有独钟呢?思索这个问题能够让我们领略图灵构想的优美和深刻。

我们知道,今天的计算机能够执行和完成许多从前被认为只有聪明的人类才能从事的任务。但是,我们大多数人不愿承认迄今为止的计算机是有智能的。鉴于这种情况,美国耶鲁大学的人工智能专家德鲁·麦克德莫特(Drew McDermott)在1997年IBM计算机深蓝击败卡斯帕罗夫后评论道:

去年[1996],在加里·卡斯帕罗夫象棋取胜IBM计算机深蓝后,我告诉我的人工智能导论课的学生说,计算机要挑战最好的人类,还需许多年。既然已经证明我和许多其他人都是错的,许多人一直急于要我们相信,深蓝并不真正具有智

能,而且这场胜利与人工智能的未来无甚干系。……尽管我同意这台计算机不是非常有智能,但是,说它根本不显示任何智能却是根本误解了它所做的事情和人工智能的目标和方法。的确,深蓝的能力非常狭窄。它不能认识、更不能拾起一枚棋子。它甚至也不能够谈谈它获胜的那局棋。既然智能的本质似乎是在种种情形下创造性的反应能力,我们难以在这一点上说这台机器有多少智能。^①

深蓝的能力是狭窄的,算题、写诗、作画的机器依然在能力上是狭窄的,我们不愿意将智力赋予它们,是因为我们认为智力是远为宽广的能力。图灵选择语言能力作为智能测试的标准,我们难道不能同样说这种能力是狭窄的吗?

在图灵看来,回答是否定的。谈话的能力在所有智力中是独特的和涵盖性的。图灵试验并不限定交谈的话题,模仿游戏的参与者可以就任何题目展开语言上的交流。如果我们将两台机器做一个比较,这一点就更容易理解。假设有一台机器,深红,它在图灵的模仿游戏中胜出。我们可以比较一下深红与深蓝的智力。深蓝只能下棋,而深红可以谈论广泛的话题(包括象棋以及深蓝的获胜局)。实际上,如果深红根本不会下棋,它也不可能聪明地谈论下棋。当然,深红在某种程度上可以被认为是“夸夸其谈”或“纸上谈兵”,如果在棋盘上与深蓝较量会败得一塌糊涂。但这并不重要,世界上没什么人在棋上能与卡斯帕罗夫相比,但他们都可以是有智能的;甚至一个完全不会下棋的人,也可以是有智能的。图灵美妙的构想的背后隐藏着这个思想,即,交谈能力不仅仅是诸多智力的一种,更为重要的是,交谈能力是一种能够(至少在某个程度上)表现其他诸多能力的的能力,是一种涵盖性的宽广的能力。套用一句话,“智能一枝花,全靠嘴当家”。

图灵在他的经典论文中做过如下预言:

我相信,在大约 50 年的时间里,人类有可能用大约 10^9

^① 德鲁·麦克德莫特(Drew McDermott):《对,计算机能思维》(Yes, Computers Can Think),《纽约时报》1997 年 5 月 14 日。

的存储量给计算机编程,让其玩模仿游戏,它们玩得如此之好,致使一个平常提问者在5分钟的提问之后做出正确的辨别的机会不足70%。……我相信本[20]世纪末,人们的说法以及一般有教育的观点将大大改观,人们将能够谈论机器思维而不感到抵触。^①

今天看,这个预言即使不能说完全失灵,至少没有得到实践上的支持。有一些哲学家和人工智能研究者从其他角度论证人工智能的局限。我把这类论证分为两类:

第一类:有一些先验证据表明,计算机系统依其本性,在模拟人类智能方面存在原则上的制约,因此,人类构造的任何计算机系统都不可能通过图灵试验;因此,人类的智能不能用计算机模型来解释;因此,机械论的智能观点是错误的。

第二类:即使计算机系统能够通过图灵试验,依其本性,它们也不具有与人类一样的智能,因此试图用计算机模型来解释人类智能的努力是无法成功的。

美国哲学家约翰·塞尔(John Searle)于1980年提出的“汉字屋”论证(Chinese Room Argument)是第二类论证的代表。^② 对于这一类论证,我在这里不作讨论。我的主要兴趣集中在第一类论证上。这类论证试图从原则考察计算机具有的能力。这里有两点值得注意。首先,人工智能是一门具体的学科,有经验探究的一面。而科学是一种构造性的工作,它需要确立何为世界中的事实以及刻画事实之间的关系。作为一门科学,人工智能的理论目标是提出一种广义的智能理论或模型,以取代大众心理学中蕴含的昏暗不清的常识智能观。哲学,就其作为最一般的概念上的工作而言,不可能代替经验的观察断言经验世界中的事实,虽然它必须谈论这些事实。哲学探讨的是各种判断之如何可能的方式,亦即确立

^① 图灵:《计算机与智能》(《Computing Machinery and Intelligence》*Mind*, vol. 59, No. 236(1950)),第442页。

^② 塞尔:《心灵、大脑与程序》(Minds, Brains, and Programs),首次发表在《行为与大脑科学》(*Behavioral and Brain Science* 1, 417~424, 1980)上。这篇文章后来被许多文集收录。对塞尔论证的评论,参见第3章“论人是机器”。