

# 认知科学导论

〔加〕 P. 萨伽德 著

朱菁 译

中国科学技术大学出版社

# 认知科学导论

P·萨伽德 著

朱 菁 译

中国科学技术大学出版社

1999·合肥

## 内 容 简 介

认知科学是对心智与智能的跨学科的研究,包括心理学、人工智能、神经科学、语言学、人类学和哲学。本书以表征-计算这一看待心智的核心理念为线索,系统地介绍和评价了当代认知科学中的不同研究路线所取得的成就,包括基于逻辑、规则、概念、类比、表象和联接(神经网络)所发展的各种认知科学理论。同时也讨论了从情绪、意识、物质环境与社会环境、动力学系统和数学知识等方面对以表征-计算观为代表的正统认知科学的挑战。本书是为各种不同学科的大学生和低年级研究生撰写的教材,内容精练,深入浅出,力求把握当前认知科学发展的最新动向。对有关领域的教师与研究人员亦很有参考价值。作者保罗·萨伽德是加拿大滑铁卢大学哲学系教授,心理学系和计算机科学系的兼职教授。

## 图书在版编目(CIP)数据

认知科学导论/(加拿大)萨伽德(Thagard, P.)著;朱菁译  
·—合肥:中国科学技术大学出版社,1999.5  
书名原文: Mind: Introduction to Coqnitve Science  
ISBN 7-312-00939-5

I. 认… II. ①萨… ②朱… III. 认知科学-概论  
IV. B842.1

中国版本图书馆 CIP 数据核字(1999)第 16522 号

中国科学技术大学出版社出版发行

(安徽省合肥市金寨路 96 号,邮编: 230026)

肥西新华印刷厂印刷

全国新华书店经销

850×1168/32 印张: 6.75 字数: 173 千

1999 年 5 月第 1 版 1999 年 5 月第 1 次印刷

印数 1—4000 册 定价: 10.00 元

**Mind: introduction to cognitive science**

Copyright © 1996, by Massachusetts Institute of Technology  
Press

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

业经授权,中国科学技术大学出版社享有本书中文简体字版专有  
出版权

著作权合同登记号:皖登字 1299025 号

## 前 言

认知科学是对心智(mind)和智能(intelligence)的跨学科的研究,包括哲学、心理学、人工智能、神经科学、语言学 and 人类学。它的学术起源在 50 年代中期,不同领域的研究者开始借助复杂的表征(representation)和计算程序来发展关于心智的理论。到 70 年代中期,其学术组织开始形成,成立了“认知科学学会”并开始出版《认知科学》杂志。至今,在北美和欧洲,已有 60 多所大学设立了认知科学专业,更多的学校则开设了认知科学的课程。

讲授认知科学这门跨学科课程有相当的难度,因为学生来自各种不同的背景。从 1993 年起,我在滑铁卢大学开设了名为“认知科学导论”的课程。一方面,这门课吸引了来自计算机科学和工程等学科的学生,他们具有较好的计算方面的背景但对心理学和哲学所知甚少;另一方面,在心理学和哲学上有良好背景的学生却对计算不甚了解。这本教材便是为建设一门毋须任何认知科学的专业准备的课程所作尝试的一个部分,其目的是使对心智和智能感兴趣的同学能够认识到在对心智的研究中有许多互补的途径。

向来自不同学科的读者介绍认知科学至少有三种方式。第一种是集中介绍各个不同的领域,如心理学、人工智能,等等;第二种是以心智的不同功能来组织讨论,如问题解决、记忆、学习和语言。我选择了第三种方式,即系

系统地介绍和评价认知科学家们提出的主要的关于心理表征 (mental representation) 的理论,包括逻辑的、规则的、概念的、类比的、表象的和联接的(神经网络)理论。对这些基本理论途径的讨论是为了以一种统一的方式来展示认知科学不同领域所取得的成就,同时注重对各种重要的心理功能的理解。

撰写此书的目的是为所有有兴趣选修认知科学导论课的学生提供一本入门书籍。为了达到这一目的,就需要以一种适当的方式向心理学系的学生介绍逻辑,向英语系的学生介绍计算机算法,向计算机科学系的学生介绍哲学上的争论。作者热情地欢迎各方面的读者提出宝贵意见。

虽然本书是针对本科生的,对于研究生和教师也有参考价值,有助于他们了解他们各自的领域在整个认知科学的事业中所处的状况。我没有试图写成一部百科全书。由于整个着眼点是提供一个整体性的导论,我既要使得全书的篇幅保持在适当的范围内,同时也力求侧重于森林而不是树木。因为将认知科学视为各个不同领域的交叉融合而不是一个统一的整体,我忽略了人工智能、认知心理学和心智哲学等导论课程里的许多标准论题。在每一章的最后都附有小结和进一步的推荐读物。每章的“备注”一节为有特别兴趣的读者提供信息。

本书强调了心理表征理论对理解心智的重要性,同时也意识到认知科学还有很长的路要走。第二部分讨论了对认知科学基本假设的挑战,并且对这一跨学科研究未来的方向提出了建议。

# 目 次

前 言	( I )
第一章 表征与计算	( 1 )
第二章 逻辑	( 19 )
第三章 规则	( 40 )
第四章 概念	( 56 )
第五章 类比	( 74 )
第六章 表象	( 91 )
第七章 联接	( 105 )
第八章 回顾与评价	( 126 )
第九章 情绪与意识	( 135 )
第十章 物质环境与社会环境	( 150 )
第十一章 动力学系统与数学知识	( 164 )
第十二章 认知科学的未来	( 178 )
附录:认知科学的资源	( 183 )
参考文献	( 186 )

# 第一章 表征与计算

## 探索心智

不知你是否曾经对于你的心智是如何工作的感到过惊奇？人们每天都要完成各式各样的心理任务：在工作和学习中解决问题，对个人生活作出决定，对所知道的人的行为给予解释，以及获取各种新的概念。认知科学的主要目的就是要去解释人们是怎样完成这各种各样的思维活动的。我们不仅要对各种问题求解和学习的过程进行描述，还要对心智是怎样去执行这些操作的给出说明。

理解心智如何工作对于许多实践活动来说都至关重要。教育工作者需要了解学生思维活动的本质，以便寻求更好的方法进行教学。工程师和其他设计人员则需要知道他们的用户在有效或无效地使用他们设计的产品时是怎么想的。通过反思是什么使人变得聪明，我们可以使得计算机变得更具智能。而政治家和决策者们如果能理解与他们打交道的人们的心理过程，他们将会取得更大的成功。

然而对心智的探索实非易事，因为我们不可能剥开一颗心智来看一看它究竟是如何工作的。千百年来，哲学家和心理学家将心智作了各式各样的比喻，例如将它比作一张白纸，在上面做各种印记；或是一种液压装置，由各种力来操作它；甚至将它比作电话开关板。近半个世纪来，更新的、富于启发性的比喻来自于各种新型计算机的发展。许多（但不是全部）认知科学家都将思维视作一种计算，并采用计算隐喻来描述和解释人们是怎样学习和求解问题的。



## 你所知道的是什么？

在同学们进入大学后，除了课程内容之外还有许多东西是需要了解的。不同专业的学生将要学习的课业千差万别，但对于大学是如何运作的都需要有所了解。如何选课？课程什么时间开始？哪些课程好而哪些课要尽量避免？要获得学位必须具备哪些条件？从一栋教学楼到另一栋哪条路径最优？校园里的其他同学是怎样的？哪儿是欢度周末的最佳场所？

这些问题的答案会成为大多数学生脑海中的一部分，但这是怎样的一部分呢？大多数的认知科学家都认为心智中的知识是由心理表征（mental representation）构成的。大家都熟悉各种非心理表征，比如说印在本页上的词语。在这里我就用了“本页”一词来代表你正在阅读的这一页。同学们会经常使用到一些形象化的表征，例如校园和教学楼的地图。为了说明像学生对大学的了解这些不同类型的知识，认知科学家们提出了各种各样的心理表征方式，如规则、概念、表象（image）和类比。学生要了解的规则有“如果我要能毕业，那么我必须选够10门本专业的课程”。他们会增加一些新的概念，例如像使用“小鸟”、“米老鼠”或是“香肠”这样一些词汇来形容一门特别容易的课。为了从一栋楼房转到另一栋楼，一幅校园布局的心理表象或图片将是十分有用的。在选修了一门特别喜欢的课程之后，可能还会选择一门类似的课程来学习。通过与校园里不同专业的同学接触，心目中可能会形成关于某种类型的学生的模板（stereotype），尽管很难准确地说出究竟是什么构成了这样一些模板。

学生掌握这样一些有关校园生活的知识并非只是为了积累信息。作为学生要面对很多的问题，诸如如何学好各自的专业，怎样拥有合宜的社会生活，以及怎样在毕业后找到工作。解决这些问题就需要与心理表征打交道，例如作出“为了能够毕业还必须

再修 5 门课”这样的推断，或是作出“千万别再选梯迭姆 (Tedium) 教授的课”这样的决定。认知科学提出人们是通过在心理表征之上运行心理程序 (mental procedure) 而产生出思维和行动，而规则和概念这些不同类型的心理表征则支持不同类型的心理程序。在此我们不妨看一下表示数目的不同方式。绝大多数人熟悉的是阿拉伯数字系统，1, 2, 3, 10, 100, 等等，以及进行加、乘等运算的标准程序。罗马数字同样可以表示数目，用 I、II、III、X、C，但要用不同的程序来进行算术运算。请试试用 X X VI (26) 去除 CV (104)。

本书第一部分将考察在认知科学研究过去的 40 年中发展起来的研究心理表征和心理程序的不同途径。关于这些不同的途径的得失已有不少的争论，许多权威的认知科学理论家为他们所偏好的研究路线热烈地争辩。我个人采取的道路较为折衷，因为我认为目前已有的不同的心理表征理论其互补性大于其竞争性。人类的心智出奇的复杂，而我们对它的理解无论从它对规则的使用 (如上述) 还是从许多其他类型的 (包括一些我们所不熟悉的) 表征方式上都将获得收益。后者包括“联接主义者” (Connectionist) 或“神经网络”式的表征，这些将在本书第七章予以讨论。

## 起 源

对心智及其活动进行理解的尝试至少可以追溯到古希腊，哲学家柏拉图和亚里士多德试图说明人类知识的本质。柏拉图认为最重要的知识来源于概念，比如美德就是这样的概念，人们不依赖于经验而能凭天赋获知。另外像笛卡尔和莱布尼茨这样的哲学家也相信只须借助于思维和推理就能获得知识，这种立场被称为唯理性主义 (rationalism)。与之相反，亚里士多德通过“所有的人都是会死的”这样的规则来探讨知识。这种哲学立场，经由洛

克、休谟和其他一些哲学家的辩护，成为众所周知的**经验主义**（empiricism）。在18世纪，康德试图调和唯理性主义和经验主义，他主张人类的知识既依赖于感觉经验，也离不开心智的天赋能力。

直到19世纪实验心理学诞生以前，对心智的研究一直停留在哲学领域。威廉·冯特及其学生开创了较为系统地研究心理过程的实验方法。然而，在其后的几十年间，实验心理学逐渐为**行为主义**（behaviorism）所统治，行为主义实质上否定心智的存在。依照像J. B. 华生（Watson 1913）这样的行为主义者的观点，心理学必须严格限于研究可观察的刺激与可观察的行为反应之间的关系，谈论意识和心理表征是为正当的科学研究所不耻的。尤其是在北美，行为主义统治心理学领域一直到本世纪50年代。

大约到了1956年，学术界的氛围开始发生剧变。乔治·米勒（Miller 1956）总结了一系列研究，表明人类的思维能力是有限制的，例如对短时记忆来说，大约就限于7个条目（这也就是为什么人们很难记住过长的电话号码和社会保障号码的原故）。他提出将信息编码分为组块（chunk）可以克服记忆容限，组块便是须要心理程序进行编码和解码的心理表征。与此同时，尽管第一代计算机才面世几年，约翰·麦卡锡、马文·明斯基、阿兰·纽威尔和赫伯特·西蒙便创立了人工智能这一研究领域。此外，诺姆·乔姆斯基（Chomsky 1957, 1959）强烈地抨击了行为主义关于语言是一种通过学习获得的习惯的主张，取而代之以由规则所构成的心理语法来说明人们理解语言的能力。上述六位思想家可以恰当地被称为**认知科学的奠基人**。

在后面的章节里我们将结合不同的心理表征理论勾画出认知科学在其后发展的历史。麦卡锡成为以形式逻辑为基础探索人工智能的领头人，我们将在第二章加以讨论。60年代，纽威尔和西蒙揭示了依据规则来说明人类智能的威力，在第三章里我们要讨论延续这一传统所进行的一些工作。到了70年代，明斯基提出类似于概念的框架（frame）是知识表达的核心形式，而人工智能和

心理学领域的其他研究者也探讨所谓程式 (schema) 和脚本 (script) 这些与之相似的结构 (第四章)。与此同时, 心理学家们开始对心理表象显示出与日俱增的兴趣 (第五章)。自 80 年代以来, 更多的实验和计算研究集中于类比推理, 也称作基于案例的推理 (第六章)。80 年代最激动人心的莫过于联接主义心理表征和加工理论的兴起, 它大体上是以前的大脑的神经网络为模型的 (第七章)。这些不同的研究途径都已对理解心智做出了应有的贡献, 在第八章将对它们的长处和缺陷作出总结和评价。尽管如此, 对于心智应当被理解为心理表征与程序这一核心论点仍存在着诸多挑战, 这些将在本书的第二部分 (第九章至第十一章) 予以讨论。

## 认知科学的方法

认知科学绝不当只是不同领域的人们在午餐会上聚在一起聊一聊心智是什么。但在我们着手审视认知科学中已取得的一致性的观念之前, 我们先得鉴别一下不同领域的研究者对心智和智能的探索所带来的视角和方法的多样性。

虽然认知心理学家今天时常介入理论和计算模型的探讨, 但他们首要的方法仍是以人为受试者进行实验。作为受试者的人, 通常是符合实验要求的本科生, 被带到实验室以便不同类型的思维过程能够在受控条件下得以研究。不妨举一些后面章节里将会提到的例子, 心理学家通过实验检验人们在进行演绎推理时所犯错误的种类, 人们形成和使用概念的方式, 人们使用心理表象完成思考时的速度, 以及人们运用类比求解问题时的表现。我们关于心智如何运作的结论绝不能仅仅建立在“常识”和内省的基础上, 因为它们提供的很可能是一幅引人误入歧途的图景, 而且许多的心理过程是意识不到的。因此, 从不同的方面慎重细致地揭示心理过程的心理学实验, 是使得认知科学成为一门科学的关键所在。

虽然理论缺乏实验是空洞的,但没有理论实验却是盲目的。为了弄清有关心智本质的根本性问题,心理学实验需要在一个以心理表征与程序为基础的理论框架中加以说明。而发展理论框架最好的办法之一是构造和测试计算模型,以便模拟心理运作。为了补充有关演绎推理、概念生成、心理表象和类比式问题求解的心理学实验,研究人员发展了各种计算模型来模拟人类完成这些活动的各个方面。设计、建造和测试这些计算模型是人工智能(AI)的核心方法,AI是研究智能系统的一个计算机科学分支。在认知科学中最理想的状况是,计算模型和心理学实验能够携手共进,但AI中很多重要的工作是在与实验心理学相对隔离的情况下检验不同的知识表达方式的能力的。

尽管有一些语言学家也做心理学实验或者研制计算模型,但大多数还是使用其他方法。对于乔姆斯基传统的语言学家来说,主要的理论任务是确定作为人类语言的基本结构的语法原则,这种鉴别任务要求对符合语法和不符合语法的言语之间的细微差别予以关注。例如,在英语中,句子“*She hit the ball*”和“*What do you like?*”是符合语法的,而“*She the hit ball*”和“*What does you like?*”则不是。英语语法就要解释为什么前者是可接受的而后者不是。在后面的章节中会给出乔姆斯基式的和其他学派的语言学家进行的理论和经验研究的例子。

同认知心理学家一样,神经科学家也进行受控实验,但他们进行的观察却大不一样,因为神经科学家直接关心大脑的本质。对非人类的受试者,研究者可以插入电极,记录单个神经元的激活。而对人类受试者,这项技术就太过于侵害性了。近年来采用磁扫描和正电子扫描装置,可以观察当人们在完成各种心理任务时大脑的不同部位所发生的情况。例如,脑扫描可以确定在运用心理表象或进行语词理解时大脑活动的区域。有关大脑功能的另一类证据来自于观察那些大脑受到某种方式的损伤后的人的行为。例如,发生中风时,大脑专管语言的一部分会导致诸如无法发出完

整句子这类的行为缺损。和认知心理学一样，神经科学既是理论性的又是实验性的，而理论的发展通常借助于研制反映神经元群组行为的计算模型。

认知人类学研究在不同的文化环境中人们如何思考，从而拓宽了对人类思维的审视视角。对心智的研究显然不应当只局限于讨论操英语的人们是如何进行思考的，而应当考虑到不同文化的思维模式可能的差别。在第十章将讨论认知科学如何越来越多地意识到需要在特定的社会环境和物质环境中考虑心智的活动。对认知人类学家来说，主要的方法是现场社群调查法，这就要求研究者同某一文化的成员共同生活，打成一片，这样才能把握他们所特有的社会和文化系统。例如，认知人类学家已经调查了不同文化中描述颜色的词语的相似和差异。

除了少数例外，哲学家一般不进行系统的经验观察或建造计算模型。但哲学对认知科学来说仍显得重要，这是因为认知科学涉及一些基本性的问题，这些问题对于心智的实验研究或计算研究的方式都是基础性的。像表征和计算的本质这种抽象问题在心理学和人工智能的日常研究中毋须顾及，但研究者对于他们的所做所为进行更深入的思考时就无法避免了。哲学还处理一些普遍性的问题，如心灵与肉体的关系，以及一些方法论问题，如认知科学中解释说明的实质。此外，哲学既涉及人们实际上是如何进行思维这种描述性问题，也关心人们应当如何进行思维这样一些规范性问题，这是因为除了理解人类思维这样的理论目标之外，认知科学还具有改善人类思维这样的实践目的，这就要求对我们想要的思维是怎样的进行规范性的反思。心智哲学本身并没有一种突出的方法，但应当与其他领域的最优秀的理论研究一同关注经验成果。

在其最弱的形式上，认知科学只是上述几个领域的汇合：心理学、人工智能、语言学、神经科学、人类学和哲学。当理论研究和实验研究在关于心智本质的结论上趋同聚合时，跨学科的研

究会变得更为引人入胜。后面的章节将会提供案例表明认知科学处于各个领域的交汇面上。从理论上讲，目前最富有成效的方式是依据表征和计算来理解心智。

## 对心智的计算-表征理解

认知科学的中心假设是：对思维最恰当的理解是将其视为心智中的表征结构以及在這些结构上进行操作的计算程序。尽管在有关构成思维的表征和计算的实质是什么这一问题上存在争议，但这一中心假说本身足以涵盖目前的认知科学对思维的理解，包括联接主义理论。为简略起见，我将基于这一中心假说对心智的理解方式称作 CRUM (Computational-Representational Understanding of Mind)。

CRUM 可能是错误的。本书第二部分将展示对这一路线的一些根本性的挑战，对于解释有关心智的一些基本事实，表征和计算可能并不恰当。但通过评价各种知识表征理论所获得的成功，我们将会看到，在对心智的理解上 CRUM 取得了可喜的进展。毫无疑问，CRUM 是至今为止在理论和实验上最为成功的方式。虽然在认知科学的诸领域中并非人人都赞同 CRUM，但从心理学和其他领域中的权威性学术期刊上我们不难看出 CRUM 是目前认知科学的正统观点。

CRUM 的成功在很大程度上得益于从计算机的发展中获得了丰富的类比。在第五章中我们会谈到，类比对于提出新的科学思想颇有助益，而将心智比作计算机对于我们理解心智提供了一种强有力的方式，远远胜过了以往提出的诸如电话开关板之类的比喻。有计算机科学背景的读者对于计算机程序由数据结构和算法组成这样的概括一定很熟悉，现代编程语言都包含一系列的数据结构，包括像“abc”这样的字符串，像 3 这样的数字，以及更

为复杂的结构，如表（A B C）和树。算法——机械式的程序——可以定义为在各种数据结构之上的操作。例如，一段名为 REVERSE 的程序可以定义为转置一个表，将（A B C）变为（C B A）。这个程序可由两个更小的子程序组成，先从一个表中取出一个元素，再把它加在另一个表的开头，使计算机先得到（A），再得到（B A），然后是（C B A），从而得到一个倒置的表。同样，CRUM 假定心智具有心理表征，类似于数据结构，而计算程序则类似于算法。图解如下：

程 序	心 智
数据结构+	心理表征+
算法	计算程序
=运行程序	=思维

这就是认知科学中正统的类比，尽管它将另一个基于大脑的类比作了一点奇异的扭曲。联接主义者提出了关于表征和计算的新颖见解，将神经元及其联接比作数据结构，而将神经元的激活和激励传播比作算法。这样 CRUM 就是处于心智、大脑和计算机之间的一个复杂的三维类比。三者中的每一个都可以用以为其他两个提供新见解。并不存在某个独一无二的关于心智的计算模型，因为不同种类的计算机和编程方法决定了心智运作方式的不同。我们大多数人今天接触的计算机是串行处理器的，一次执行一条指令，但是大脑和某些最新开发的计算机是并行处理器的，可以一次做很多个操作。

如果你对计算机有相当的了解，从计算机来理解心智就来得相当自然，即便你不一定同意心智从根本上说是一台计算机。从未编写过计算机程序但使用过食谱的读者可以考虑另一个类比。一个食谱通常有两个部分：一张调料表和一套调配调料的指令集。一盘菜是将烹调指令运用于调料的结果，恰如一段可执行的程序



是将算法运用于诸如数字和表这样的数据结构的结果。将菜谱来比思维有些牵强，因为调料不是表征，而烹调指令需要有人对其进行解释。第二章至第七章会提供一些更为直接对应于心智操作的计算程序的例子。

## 理论、模型和程序

对于心理过程的理论研究来说，计算机模型通常是十分有用的。对认知科学模型的把握要注意四个方面的区别与联系：理论、模型、程序和平台。一个认知理论要假定一套表征结构和一套在这些结构上进行操作的加工过程。通过与计算机程序由数据结构和算法构成进行类比说明，一个计算模型使得这些表征结构和过程更为精确。有关表征的模糊概念由准确的关于数据结构的计算概念加以补充，而心理过程则可由算法来定义。为了测试该模型，必须用一种编程语言（比如 LISP 或 C）将其在一个软件程序中实现。此程序可以在各种硬件平台上运行，例如 Macintosh、SUN 工作站或 IBM PC，或是某种专门设计的专用硬件设备，如一台带有多个并行处理器的 Connection Machine。各种结构和加工过程均可依此方法进行研究，从一些传统人工智能的规则和搜索策略，到较新的联接主义的分布式表征和传播激励加工。

举例来说，例如要了解儿童怎样学习把两个数相加，如  $13+28$  这样的问题。一个认知理论就需要指出儿童如何表征这些数字以及怎样加工这些表征以完成加法运算。这个理论要指出 13 是由单一结构来表示，还是由诸如 10 加上 3 这种复合结构，或是一种类似神经元的复杂结构来表示的。该理论还要指出在这些结构上操作以得到结果 41 是怎样一些过程，包括设法将 30 加上 11 转变成 41 的移位操作。一个计算模型通过刻画可编程的结构和算法从而更为详尽、准确地说明表征和加工处理的实质，使之与做加法的心理表征和加工过程相类似。为了评估该理论和模型，我们