

1

Nils J. Nilsson

2

3

Understanding
Beliefs

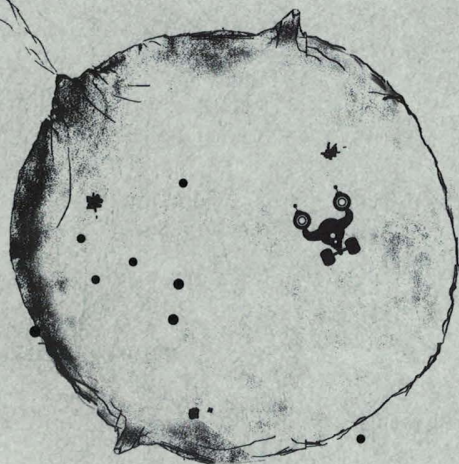
Understanding Beliefs

理解信念

人工智能的科学理解

[美] 尼尔斯·尼尔森 (Nils J. Nilsson) —— 著

王飞跃 赵学亮 —— 译

机械工业出版社
CHINA MACHINE PRESS

信念构成了我们对世界认知的很大一部分。我们不但有关于物体、关于文化、关于过去和关于未来的信念，而且有关于他人的信念，并且相信他们也有信念。我们利用信念去预测、去解释、去创造、去安慰、去娱乐，并把其中的一些信念称之为理论，我们在构建它们时极具创造力。量子力学理论、进化论和相对论便是例子。这些理论都是我们丰富的思想试图为观察到的现象寻找解释的产物。在本书中，尼尔斯·尼尔森教授将揭示我们的信念能为我们做什么、我们是如何持有它们的以及如何评价它们的。

Nils J. Nilsson

Understanding Beliefs

理解信念

人工智能的科学理解

【英】尼尔斯·尼尔森——著
王飞跃 赵学亮——译

Nils J. Nilsson

Understanding Beliefs

理解信念

人工智能的科学理解

【英】尼尔斯·尼尔森——著

王飞跃 赵学亮——译

我们的信念构成了我们对世界的认识的很大一部分。作者通过本书带我们一起检查信念，了解信念为我们做什么，我们是如何持有信念的，以及如何评价信念。作者指出，应该仔细评价我们的信念，因为它们影响着我们如此多的行动和决定。我们所持有的信念可通过概率量化，并通过信念网络影响其他信念的概率。故此，我们可以通过调整科学方法的一些做法，并通过咨询专家的意见，来重新评价我们的信念。

Understanding Beliefs /The MIT Press/1 - 1/by Nils J. Nilsson / 978-0-26-252643-2
Original English Language Edition published by Massachusetts Institute of Technology/
Copyright © 2014 Massachusetts Institute of Technology. All Rights Reserved.

This title is published in China by China Machine Press with license from Massachusetts Institute of Technology. This edition is authorized for sale in China only, excluding Hong Kong SAR, Macao SAR and Taiwan. Unauthorized export of this edition is a violation of the Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书由 Massachusetts Institute of Technology 授权机械工业出版社在中华人民共和国境内（不包括香港、澳门特别行政区及台湾地区）出版与发行。未经许可之出口，视为违反著作权法，将受法律之制裁。

北京市版权局著作权合同登记 图字：01 - 2016 - 3778 号

图书在版编目（CIP）数据

理解信念：人工智能的科学理解/（美）尼尔斯·尼尔森（Nils J. Nilsson）著；王飞跃，赵学亮译。—北京：机械工业出版社，2016.11

书名原文：Understanding Beliefs

ISBN 978 - 7 - 111 - 55213 - 0

I. ①理… II. ①尼… ②王… ③赵… III. ①人工智能-研究 IV. ①TP18

中国版本图书馆 CIP 数据核字（2016）第 248850 号

机械工业出版社（北京市百万庄大街 22 号 邮政编码 100037）

策划编辑：坚喜斌 责任编辑：於薇 刘林澎 杨冰

责任校对：赵蕊 版式设计：张文贵

涿州市京南印刷厂印刷

2017 年 4 月第 1 版·第 1 次印刷

145mm×210mm·6.625 印张·3 插页·68 千字

标准书号：ISBN 978 - 7 - 111 - 55213 - 0

定价：39.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务

网络服务

服务咨询热线：(010) 88361066

机工官网：www.cmpbook.com

读者购书热线：(010) 68326294

机工官博：weibo.com/cmp1952

(010) 88379203

教育服务网：www.cmpedu.com

封面无防伪标均为盗版

金书网：www.golden-book.com

名言警句

[信念] ……一个人遵照其行动。

——亚历山大·贝恩 (Alexander Bain), 心理学家,
引自路易斯·梅南 (Louis Menand),
《哲学俱乐部》¹ (*The Metaphysical Club*)

人最有价值的特质是有一种不相信什么的明智意识。

——欧里庇得斯 (Euripides),
约公元前 480 年—前 406 年

我们可能不知道很多,但确实知道一些东西,而且我们必须随时准备改变想法,我们必须依据我们确实知道的,尽我们所能地行动。

——W·H·奥登 (W. H. Auden),
引自亚当·高普尼克 (Adam Gopnik),
《双面人》(*The Double Man*)²

中文版作者序

王飞跃教授和赵学亮博士将我的书《Understanding Beliefs》翻译成中文是我的荣幸。为什么我觉得中国读者可能会对这本书感兴趣？中国人有理由为他们的古老文化传统以其许多受人尊重的和深厚的信念而感到自豪。然而，正如我在此书的英文版序言中写到的那样，“因为信念影响我们的行动，更因为我们的一些行动可能会产生严重的后果，所以我认为周密地评价信念是十分重要的”。——即使是古老而受人尊重的信念也是如此。

1981 年，我来到中国做了关于人工智能的讲座。那趟为期三周的旅行让我对中国文化及其传统印象深刻。2007 年，我和妻子再次来到中国——那次有四个

理 解 信 念
人工智能的科学理解

星期。我们感受到中国正在成为一个繁荣的现代化国家。就像我常向王飞跃教授提到的：我的希望是，中国和西方国家将共同合作建设一个和平而繁荣的世界。

感谢王飞跃教授的努力，特别是他和张钹教授为本书所撰写的序言。希望读者发现本书有益，理解信念，享受生活。

尼尔斯·尼尔森

推 荐 序

信念，《现代汉语词典》解释为“自己认为可以确信的看法”。据此定义可以看出，信念具有很强的主观性。人人都会有信念，但未必都考虑过信念本身的问题。本书中，尼尔斯·尼尔森教授结合自身经历，探讨了信念的产生、形成、性质、作用、影响及其评价方法，讲述了他的元信念，即他自己关于信念的信念。全书涉及科学、哲学、社会学及心理学，内容丰富，观点新颖，耐人寻味。

书中提到，人们主要通过感觉，如看、听、触、读等，给所相信的事物建立解释以及衍生结果来获取信念。而感知到的东西主要取决于人们基于信念的期望，因此人们时常“看到”所期望看到的事物，而看

不到不期望的，所以眼见不一定为实。人们应该用判断性思维检验信念，而判断性思维的要素包括寻求专业人士的观点，考虑信念的解释与结果，并消除能被更可信的方案替代的解释。依据统计学，信念的强度可使用概率定量表示，即通过频率方法或主观概率估计确定。信念可在一种“贝叶斯信念网络”中进行计算，该网络中每个信念的可信度都会影响其他信念的可信度。本书中详细介绍了科学方法、科学知识的本质和科学探索的基本过程，基于科学的各种实验方法也为评价信念提供了可行的参考方案。尼尔森认为，所有存储在一个计算机系统内的陈述性信息就构成它的信念，因此机器人也有信念，如谷歌的无人驾驶车和 IBM 的沃森。而且，在形成有用信念的能力上，机器人与人类几乎处在同一水平上。尼尔森强调，人们的日常信念应该接受变化，如同科学理论的发展一样，应逐步完善。由于心理因素及生活方式的原因，人们

推荐序

容易被一些“信念陷阱”所困，而应对陷阱的最好方法，就是将信念置于他人“理由充分”的批评之下。

尼尔斯·尼尔森是人工智能学科的奠基性研究者之一，他从个人的研究背景出发，以独特的视角解读信念的内涵，论述深刻，说服力强，值得一读。

张 钹

清华大学教授，中国科学院院士

译者序

科学与信念

一年多前，我去俄勒冈州拜访在梅德福享受退休生活的斯坦福大学计算机系前主任、人工智能实验室的创始元老尼尔斯·尼尔森教授。聊天时，望着从自己后院流过的清清小溪，尼尔森突然对我说，他越来越不相信绝对真理的存在，真理不过是非常合理而且十分可靠的信念而已！为此，他专门写了本小册子，就是刚由麻省理工学院出版社出版的《Understanding Beliefs》《理解信念：人工智能的科学理解》。

我一时无语，感到愕然。这似乎与我“认识”的尼尔森不同。我从来都认为尼尔森是推动人工智能研

究从“文学”到科学的最大功臣，一生追求科学“真理”，治学极其严谨。我本人也曾深受其影响，30多年前通读了他在机器学习、人工智能和智能机器人方面几乎全部的著作和论文，还写过数篇评论文章。尼尔斯曾是以逻辑推理为主导的人工智能“纯净派”（Neats）的先锋大将[⊖]，就连实质为连续逻辑的模糊逻辑都被其斥为“暂时的癖好”（temporary idiosyncrasy）[⊙]。因此，他一直给人以一种“要么绝对是真理，要么什么都不是”的绝对“正统”的感觉。记得1992年我本人带学生去加州硅谷圣何塞参加第一届AAAI（当时的美国人工智能学会，现在的人工智能

⊖ 王飞跃. 建立人工智能的数学体系——介绍《Logical Foundations of Artificial Intelligence》[J]. 计算机科学, 1989 (2): 79-80.

⊙ Nils J. Nilsson. On Logical Foundations of Artificial Intelligence: A Response to the Reviews by S. Smoliar and J. Sowa, Artificial Intelligence, 1989, 38(1): 132-133.

促进会)举办的机器人比赛和展示时,与时任 AAAI 主席的尼尔斯不期而遇。当时,他对智能机器人的未来发展所持的态度也给我一种基于逻辑之必然“真理”的感觉。眼下,尼尔斯对真理的“新”认识,多少有些让我感觉是对“真理”的“模糊逻辑”式的理解,但我相信这不是尼尔斯“暂时的癖好”。

其实,尼尔斯现在对真理的看法更能引起我的共鸣。关于真理,从来都有客观与主观、唯物与唯心的争论。叔本华把作为真理载体的世界视为意志和表象的合成,尽管不同哲学家对谁为意志、谁为表象的看法往往截然相反,但多数认可其源自柏拉图的“理念”是意志和表象之间的一个中介的观点。或许,信念就是多数人“相信了的理念”的简称,是人们进行思维活动的工具和表达思维活动结果的方式而已。

问题是,思维本身也是时间或时代的函数。思维是人类认识世界的方式和手段。狭义地讲,就是人们

思考问题、表述问题、讨论问题的“框架”(Framework)。在不同的历史阶段,人们对思维“框架”的认识也不同,古希腊的柏拉图有“蜡板”之说,工业革命前夕的法国哲学家笛卡儿有“天赋观念”之论,英国哲学家洛克还有“白板”理论,等等。正如后来德国哲学家恩格斯所言:“每一个时代的理论思维,包括我们这个时代的理论思维,都是一种历史的产物,它在不同的时代具有完全不同的形式,同时具有完全不同的内容”。^①

我们处于什么时代?我们又需要什么样的思维和信念?这是一个重要的问题。我认为人类社会已进入开发波普尔的第三世界“人工世界”(虚理世界)的

① 马克思恩格斯选集(第3卷)[M]. 北京:人民出版社, 1972: 465.

“新 IT”（Intelligent Technology，智能技术）时代，我们需要“计算思维”和“智能信念”^{①②}。对此，作为人工智能和智能机器人的杰出开拓者，尼尔斯的《理解信念：人工智能的科学理解》一书为我们提供了很好的思考材料和启示路径。

在文艺复兴和科学兴起之前，人类追求的主要是神圣的“上帝认可”，期盼“天人合一”，否则心中不安。现代科学体系确立之后，我们的追求变为牛顿定律般永恒的因果律“理性认可”，希望“知行合一”，否则就不彻底。今天，物联网、大数据、云计算等新 IT 技术唤醒了人们对个性化、针对性、瞬变态等“长

① 王飞跃. 从计算思维到计算文化 [J]. 中国计算机学会通讯, 2007, 3 (11).

② 王飞跃. “阿尔法围棋”走向何方 [J]. 光明日报, 2016 年 07 月 08 日 10 版.