

• 科学技术哲学 •

从物理学到人工智能

——2024年诺贝尔物理学奖引发的争议及其解析

From Physics to Artificial Intelligence:

An In-depth Analysis of the Controversies Triggered by the Nobel Prize in Physics 2024

朱菁 / ZHU Jing¹ 詹好 / ZHAN Hao²

(1. 厦门大学哲学系, 福建厦门, 361005; 2. 山西大学哲学学院, 山西太原, 030006)

(1. Department of Philosophy, Xiamen University, Xiamen, Fujian, 361005;

2. School of Philosophy, Shanxi University, Taiyuan, Shanxi, 030006)

摘要: 2024年诺贝尔物理学奖授予了一位由物理学家转行的分子生物学家和一位从事人工智能研究的计算机科学家, 以表彰他们为基于人工神经网络的机器学习所做出的奠基性贡献。这一决定引发了广泛的关注、质疑和争议, 公众普遍对获奖成就的学科归属感到疑惑。鲜为人知的另一个争议是机器学习的业内专家批评此决定损害了科学的严正性, 是变相奖励剽窃以及不正当的荣誉归属。本文对这两大争议进行深度解析, 揭示重大科研成果的首创权认定和成就归属过程中不容忽视的复杂社会因素, 并提出关于科学发现首创权的认定和荣誉分配问题的新见解。

关键词: 科研奖励 物理学 人工智能 神经网络 跨学科研究

Abstract: The Nobel Prize in Physics 2024 was awarded jointly to John Hopfield, a physicist-turned molecular biologist and computational neuroscientist, and Geoffrey Hinton, a computer scientist for his whole career, “for foundational discoveries and inventions that enable machine learning with artificial neural networks.” This decision has triggered widespread confusion and controversy in the public in regard to awarding a top prize in physics to achievements in artificial intelligence. This reflects a lack of knowledge among the public of the evolution of disciplinary boundaries caused by interdisciplinary research originating in physics. Another cluster of disputes and criticisms, is related to the charge made by the computer scientist Jürgen Schmidhuber that the Nobel Prize in Physics 2024 is a Nobel Prize for plagiarism, for they republished methodologies and techniques developed earlier for artificial neural networks without citing the original works. This article provides an in-depth analysis of these controversies, with a focus on the complicated social contexts that underlie the complex development from neural networks to deep learning inspired by ideas and techniques from physics, and discusses the sophisticated relations between priorities in scientific discovery, the reward system in science and the appropriate allocation of credit and honors, in light of insights from the sociology of science and recent developments in scientometrics.

Key Words: Scientific rewards; Physics; Artificial intelligence; Neural networks; Interdisciplinarity

中图分类号: TP18; C1 DOI: 10.15994/j.1000-0763.2026.11.004 CSTR: 32281.14.jdn.2026.11.004

收稿日期: 2025年4月26日

作者简介: 朱菁 (1969-) 男, 湖南湘潭人, 厦门大学哲学系教授, 研究方向为科技哲学、知识论、分析哲学等。Email: zhujing@xmu.edu.cn

詹好 (1993-) 男, 四川泸州人, 山西大学哲学学院讲师, 研究方向为人工智能哲学、游戏哲学。Email: zhanhao@sxu.edu.cn

引言

2024年10月8日,瑞典皇家科学院宣布将当年的诺贝尔物理学奖授予约翰·霍普菲尔德(John Hopfield)和杰弗里·辛顿(Geoffrey Hinton),以表彰他们所作出的“奠基性的发现和发明,使得基于人工神经网络的机器学习能够展现威力”。这一出人意料的颁奖决定即刻引发了舆论大潮,除了普遍的惊愕和不解,也不乏质疑和非议。这是物理学奖首次授予非(传统)物理学的重大成果,且获奖者之一(辛顿)的身份并非物理学家,而另一位获奖者(霍普菲尔德)虽然是物理学家出身,但早在1980年就已转行生物学和神经科学。伦敦帝国理工学院的一位天体物理学家在其社交媒体个人账号上写道:“我很无语。我和其他人一样喜欢机器学习和人工神经网络,但却很难视之为物理学的发现。我认为诺贝尔评奖委员会是着了人工智能炒作的道”。^[1]一位物理学专业出身的网红科普作家也评论道:霍普菲尔德和辛顿的研究是属于计算机科学领域的,将物理学奖授予他们,遮掩了物理学的光芒。连获奖者之一的辛顿在接受纽约时报采访时也表示:如果有诺贝尔计算机科学奖,那才算是名副其实。^[2]英国计算机科学家、联合国的人工智能高级顾问温迪·霍尔(Wendy Hall)女爵士对路透社记者说:“诺贝尔评奖委员会不希望错过当今的人工智能浪潮,所以他们富有创意地把杰弗里(辛顿)塞入物理学赛道。我想说把物理学奖授予这两位获奖者似有可疑之处,但他们在科学上的贡献值得获得诺贝尔奖,除此之外,还能有什么方式来嘉奖他们呢?”^[3]

有不少科学家对2024年诺贝尔物理学奖的归属表示赞同。尽管基于人工神经网络的机

器学习并非传统意义上的物理学,但有不少基本概念和技术源于物理学,物理学对于发展和理解人工智能具有重要价值。况且,物理学本身的边界也是在变化的,2021年诺贝尔物理学奖得主乔治·帕里西(Giorgio Parisi)说道:“物理学的范围正在变得越来越宽广,涵盖了许多以往并不存在的知识领域,或者在过去不属于物理学的领域。我认为诺贝尔物理学奖应当继续关照物理知识的新领域”。^[1]2004年诺贝尔物理学奖获得者弗兰克·维尔切克(Frank Wilczek)应邀为“墨子沙龙”微信公众号撰写专栏文章,对2024年诺贝尔物理学奖提供了有深度的解读,认为“这次的选择是恰当的,甚至是富有启示性的”。^[4]

在公众视野之外,还有针对此次物理学奖人选的另一种质疑和非议。德国计算机科学家尤尔根·施密德胡伯(Jürgen Schmidhuber)于颁奖决定公布的次日在其个人社交媒体账户发文,批评两位获奖者未曾正确引用和恰当承认日本科学家甘利俊一(Shun'ichi Amari)和乌克兰科学家阿列克谢·伊瓦克年科(Alexey Ivakhnenko)等人的先驱性工作,甚至有剽窃之嫌,而诺贝尔奖于是成为变相奖励剽窃和不正当的荣誉归属^①。这一指责尽管未引起公共媒体的关注,但在专业圈内还是引发了一定范围的议论,本次诺贝尔物理学奖这一最高科学大奖非同寻常的颁授,反映出科学共同体在重要成果的首创权认定、科学贡献的承认以及名誉分配等经典科学社会学议题上所呈现的新形态。

2024年诺贝尔物理学奖的获奖人选及其成就认定确实非同寻常,随后陆续出现了数量可观的解读性文献,多出自物理学家之手,以阐明这一颁奖决定虽则看似牵强附会,实则是理所当然、顺理成章的^②。物理学的思想、概念、

①参见施密德胡伯于2024年10月9日在X(2023年7月以前为推特(Twitter))的个人账号发布的评论,其网址为:<https://x.com/SchmidhuberAI/status/1844022724328394780>。

②此处仅列举部分主要由物理学家撰写的中文文章。赵鸿、张勇:“开疆拓土、再创物理辉煌——2024年诺贝尔物理学奖解读”,中国科学杂志社微信公众号,2024年10月21日;鄂维南:“2024年诺贝尔物理学奖评述”,《中国科学基金》,2024年第6期;黄海平:“神经网络研究的物理学思想”,《现代物理知识》,2024年第6期;唐乾元:“物理学与人工智能的连接:2024年诺贝尔物理学奖解析”,《科学通报》,2025年第10期。

模型等对于人工神经网络和机器学习的发展产生过重要的作用,这是以往对于人工智能及其发展史的认知所忽视的,而将诺贝尔物理学奖颁授给这方面的奠基性工作,无疑有助于矫正先前的偏颇。^{[5], [6]}此外,公众对于物理学的刻板印象也是妨碍其理解和接受2024年物理学奖的原因之一。百余年来,物理学始终是一个生机勃勃的领域,其疆域不断在扩张,人工智能和神经网络等已然成为物理学家的研究对象。不过,应景式的解读大多侧重于探讨物理学的概念、模型和方法等对人工智能的影响,鲜有对历史发展脉络和复杂社会因素的揭示和分析,本文力求弥补此方面的不足,着重探究在成就2024年物理学奖的历程中至关重要的社会因素。

相比之下,关于剽窃嫌疑以及荣誉归属的问题只是在较小的专业圈子内引起过一定的讨论,深入的相关研究和分析尚未出现。这项指责并非空穴来风,类似的情况也有先例,例如施密德胡伯曾公开指责辛顿、杨立昆(Yann LeCun)和约书亚·本吉奥(Yoshua Bengio)“三人团”多次剽窃他的成果,欺世盗名,却被誉为深度学习之父,共享了2018年的图灵奖。^[7]通过文献引用和其他方式承认他人科研工作的贡献和首创权,是科学界通行的行为规范;科研奖励尤其是诺贝尔科学奖和图灵奖等顶级大奖,是对科学家贡献和成就至高无上的承认和荣誉。由此引发的争端,历来是科技史、科学社会学、科学哲学以及科学技术与社会等学科感兴趣的问题。本文尝试对2024年物理学奖引发的相关问题进行深入的探讨,引入关于科学发现和发明的“首创权”的社会学视角来进行分析,并针对科学奖励和荣誉分配的机制提供一些新的看法和见解^①。

一、两位获奖者的非凡人生

荣获2024年诺贝尔物理学奖的两位科学家

均具有鲜明的跨学科背景,其成长经历和职业生涯非同寻常,不拘一格。了解他们各自的非凡人生,对于评价他们杰出的科学贡献以及荣膺诺贝尔物理学奖资格,无疑是有必要的。

1. 霍普菲尔德

霍普菲尔德于1933年出生在美国,父母均为物理学家。他于1954年毕业于斯沃士莫尔学院,主修物理学。1958年在康奈尔大学获得博士学位,研究领域为固体物理学。霍普菲尔德随后加盟了贝尔实验室,两年后前往巴黎高等师范学校担任研究科学家。1961年他获得了加州大学伯克利分校物理系教职,三年后成为普林斯顿大学物理系教授。作为一名物理学家,霍普菲尔德在凝聚态物理学领域取得了突出成就,1969年与贝尔实验室的老同事化学家大卫·托马斯(David G. Thomas)因关于化合物半导体和半导体光学的理论和实验研究,共享了由美国物理学会颁发的巴克莱奖(Oliver E. Buckley Prize),这是凝聚态物理学的最重要奖项。霍普菲尔德分别于1973和1976年当选了美国国家科学院和美国文理科学院院士。

1969年霍普菲尔德获得古根海姆项目的资助前往剑桥大学卡文迪许实验室访学半年。在50年后撰写的具有学术自传性质的文章中,他回忆说当时他感到在凝聚态物理中找不到适合他才智和兴趣的问题,前往剑桥的目的是希望找寻新的研究方向,却一无所获。^[8]之后不久,在为贝尔实验室半导体研究组提供咨询的工作中,霍普菲尔德遇到了化学家罗伯特·舒尔曼(Robert Shulman),后者彼时正在对血红蛋白开展高分辨率核磁共振的实验研究。舒尔曼成功地激发了霍普菲尔德的兴趣并加入其研究小组,由此开启了后者从物理学向生物物理和生物化学领域的旅程。首先是从凝聚态物理学转入生物物理和生物化学领域,霍普菲尔德于1974年提出了动力学校对(Kinetic proofreading)的概念,这是一种修复生物化学反应错误的机制,可用以解释DNA复制的准确性。^[9]1977年他受到

①此处的“首创权”对应的英文是priority,相关的中文文献历来译为“优先权”。本文采用“首创权”,以避免“优先权”所蕴含的特权意味。

麻省理工学院的神经科学家弗朗西斯·施密特（Francis Schmitt）的邀请，加入了其创立并主持的神经科学研究项目，这是一个高水平的跨学科研究群体，涉及的核心领域包括分子生物学、神经科学和心理学等。得益于一些机缘巧合，霍普菲尔德于1980年成为加州理工学院化学系和生物学系的双聘教授，实现了从物理学到分子生物学和神经科学的华丽转身。在十年的转型期中，霍普菲尔德始终在寻找能够匹配他志趣和才能的重大科学问题。

入职加州理工学院后不久，霍普菲尔德与物理学家理查德·费曼（Richard Feynman）以及微电子和计算机科学家卡维尔·米德（Carver Mead）共同开设了一门名为“计算之物理学”的本科生课程。三年之后，这门课演变为“计算与神经系统项目”的跨学科研究生项目。由霍普菲尔德担任首届主任。该项目汇集了多个不同学科的教师，已培养了100多名博士毕业生，是非常成功的计算神经科学的教学与科研平台。1997年霍普菲尔德重返普林斯顿大学，担任分子生物学系教授，帮助创立了该校的神经科学研究所（成立于2004年）；他于2008年退休，成为该校的荣休教授。

纵观霍普菲尔德的职业生涯，可以大体上分为两个阶段：1980年以前其主要身份是物理学家，之后则是分子生物学家和神经科学家。物理学奠定和塑造了他的学科底色和思维方式，但他一直在努力探寻最能够发挥其优势和专长的重大科学问题，勇于突破既有的学科边界，在物理学、化学、神经科学和分子生物学等多个领域均做出了杰出贡献。在获得2024年诺贝尔物理学奖之前，霍普菲尔德已多次获得重要科学奖项：1983年获麦克阿瑟奖，1985年获美国物理学会颁发的生物物理奖，1988年获迈克尔逊-莫雷奖，1997年获国际电气电子工程师学会颁发的神经网络先驱奖，1999年获国际神经网络学会颁发的霍姆赫兹奖，2001年获国际理论物理中心颁发的狄拉克奖，2005年获世界文化理事会颁发的艾尔伯特·爱因斯坦奖，2009年获国际电气电子工程师学会颁发的弗兰克·罗森布拉特奖，2012年获国际神经科学学

会颁发的斯瓦芝奖，2019年获本杰明·富兰克林物理学奖，2022和2023年接连斩获国际纯粹与应用物理学联盟颁发的玻尔兹曼奖（统计物理）和生命物理学奖。这些荣誉当中有超过半数是物理学领域的重要奖项。

2. 辛顿

辛顿于1947年出生在英国，其父系和母系家族均有显赫的学术世家传统，产生过包括布尔代数的创立者乔治·布尔在内的一大批学者，其父是有名的昆虫学家，入选了英国皇家学会。辛顿在剑桥大学完成了本科教育，期间他一直在探寻最适合自己的学科领域。他先是选择了物理学与生理学，希望成为一名生物物理学家，但不久后就发现他并不胜任物理学专业；此后他尝试学习哲学，一年后发现自己并不喜欢；最终他选择了心理学，尽管也不甚合意。^[10]毕业后他没有选择直接升学，而是做了一段时间的木匠学徒。1972年辛顿进入爱丁堡大学攻读博士学位，拜师克里斯托夫·朗古伊特-辛吉斯（Christopher Longuet-Higgins）。朗古伊特-辛吉斯原本是一位杰出的理论物理学家和化学家，曾在剑桥大学任教多年，1966年来到爱丁堡大学参与创建名为“机器智能与知觉”的新机构，从而转向人工智能和认知科学研究。辛顿之所以投身朗古伊特-辛吉斯门下，本意是想通过神经网络研究人工智能，而这一想法源于高中时期一位同学的启发，彼时他们讨论过大脑的工作原理或许具有全息图的特征，而朗古伊特-辛吉斯在1970年前后曾与学生合作研究过基于神经网络构建具有全息图特征的联想记忆模型。不巧的是，在辛顿入学时，朗古伊特-辛吉斯已改变了研究方向，不再认为神经网络是研究智能的正确路径，而是转向了当时主流的基于符号和语言的传统路线。师徒二人因此产生了根本性的分歧，但辛顿还是坚持自己的想法。1974年朗古伊特-辛吉斯前往苏萨克斯大学任职，辛顿也跟随前往学习，最终于1978年以研究计算机视觉的论文获得了爱丁堡大学的人工智能博士学位。

1978-1980年辛顿在美国加州大学圣迭戈分校做了两年的博士后，这对于他日后的学

术发展而言极为关键。1980年代的圣迭戈分校是认知科学的重要策源地,具有浓厚的跨学科和多元化的研究氛围,而彼时对于神经网络的研究又开始回到议事日程中,这对于辛顿而言可谓恰逢其时,躬逢其盛。1979年辛顿与布朗大学的认知科学家詹姆斯·安德森(James Anderson)共同组织了“联想记忆的并行模型工作坊”,这次会议对于辛顿来说具有特别重要的意义。^[11]其一是经历了十几年的沉寂,基于神经网络的认知科学和人工智能研究者通过这次会议开始集结,尤其是吸引了一批新生代学者,初步形成了具有共同旨趣和取向的学术共同体,开启了神经网络研究和联结主义在1980年代中后期复兴的序幕。其二是辛顿在这次会议上结识了当时在哈佛大学医学院做博士后的特伦斯·谢诺夫斯基(Terrence Sejnowski),后者不久前在普林斯顿大学获得物理学博士学位,其导师正是霍普菲尔德。谢诺夫斯基在就读博士期间,兴趣从物理学转到生物学和神经科学,与导师霍普菲尔德的转向可谓不谋而合。辛顿与谢诺夫斯基随后成为密切的合作者以及终身的挚友和学术同盟军。辛顿在加州大学圣迭戈分校的另一重大收获是作为团队青年骨干参与了并行分布式处理小组的活动,这是由心理学家大卫·罗姆哈特(David Rumelhart,他是辛顿的博士后合作导师之一)和詹姆斯·麦克兰德(James McClelland)领衔的合作团队,共有十几位成员。这个项目旨在摒弃当时认知科学和人工智能基于符号、规则和逻辑的主流范式,而是探索由类似于神经元的简单单元所构成的网络所具有的信息加工模式的研究进路。作为青年骨干,辛顿对于这个项目的发起和推进起到了关键性的作用,团队最终于1986年出版了两卷本的著作,在当时产生了现象级的广泛影响,迎来了神经网络研究在1980年代中后期的大复兴。^[12]国际神经网络学会成立于1987年,不到两年的时间其成员数超过了3500人(当今的会员数为1500余人);首届国际神经网络学术会议于当年在圣迭戈举

行,有近2000人参加,首届神经信息处理系统(Neural Information Processing Systems)年会于同年底举办;国际期刊*Neural Networks*于1988年正式出版。

在重返位于英国剑桥的医学研究委员会应用心理学部工作了两年之后,1982年辛顿加盟了美国卡耐基-梅隆大学计算机科学系。因不满当时美国里根政府治下的政治氛围以及美国人工智能研究过于依赖军方资助的状况,辛顿于1987年前往加拿大,担任多伦多大学计算机科学系教授和加拿大高等研究院(Canadian Institute for Advanced Research)研究员。1998年辛顿到伦敦大学学院创建了盖茨比计算神经科学研究所,但三年后又重回多伦多大学,直至2014年初成为荣休教授。2013年辛顿加盟了谷歌公司担任技术领导,直到2023年5月辞职,以便能够自由地谈论人工智能带来的风险。2017年他参与创建加拿大的人工智能向量研究院(Vector Institute for Artificial Intelligence)并担任首席科学顾问。辛顿是加拿大皇家学会、英国皇家学会、美国文理科学院、美国国家工程院和美国国家科学院等学术荣誉机构的会士或外籍院士,获得过包括图灵奖在内的多种最高级别的学术奖励,被誉为“深度学习之父”和当今的“人工智能教父”。

二、从霍普菲尔德网络到深度学习

在诺贝尔奖的官方网站专属于2024年物理学奖的网页上,提供了三份介绍和解读性的文字材料,可分别称之为新闻版、大众版和学术版^①。三个版本因面向不同的读者群体,在内容的侧重和取舍上有所不同,但叙事的要点基本上一致:人工神经网络研究源于对大脑和神经元结构和信息加工的模拟和建模,肇始于1940年代,但在1960年代后期遭遇了寒冬期。霍普菲尔德在1982年提出了一种能够实现联想记忆的人工神经网络,这一突破性的工作具有重要的历史意义,其中运用物理学的概念和理

①参见相关网址:<https://www.nobelprize.org/prizes/physics/2024/summary/>。

论来刻画网络的结构和特征令人耳目一新。辛顿与谢诺夫斯基从霍普菲尔德网络出发,引入19世纪的物理学家玻尔兹曼的统计物理学理论,提出了一种新的神经网络模型:玻尔兹曼机。在1990年代后期人工神经网络研究再次陷入低潮期时,辛顿仍坚持不懈,提出了改进型的模型:受限的玻尔兹曼机,在算法和学习效率上有了较大提高。2006年辛顿及其合作者依托受限的玻尔兹曼机推出了具有多个隐藏单元层的神经网络模型以及预训练方法,可用于图像识别、语言处理等任务。^[13]霍普菲尔德和辛顿自1980年代以来的一系列工作为发生在2010年前后的机器学习革命奠定了基础。

此后的各种解读均以上述叙事框架为蓝本,侧重于从技术层面阐明物理学的理论资源是如何促进了神经网络和机器学习研究的发展。与之不同,本文侧重于探讨和分析上述叙事的社会因素。

1. 霍普菲尔德网络

霍普菲尔德于1982年在《美国科学院院刊》发表的论文无疑是具有划时代意义的重大成果,也是最终导致他获得诺贝尔奖的主要成就。^[14]此文提出了一种所有神经元具有对称性全联结的神经网络模型,并以此来实现联想记忆的认知功能,创新性地运用源自统计物理学和动力学系统的概念和模型对网络的主要特征加以刻画和解释,并进行了计算模拟研究。论文发表后很快就引起了关注,两年后就有越来越多的文献将文中的神经网络冠名为“霍普菲尔德网络”和“霍普菲尔德模型”。尤为罕见的是,此文自发表以来一直保持着高频次的年度引用量,经久不衰,目前谷歌学术收录统计的引用数已接近3万次。

多年以后,霍普菲尔德追忆了这篇论文的写作和发表过程。^[8]他是从1981年初开始研究联想记忆,运用统计物理学关于磁性的自旋系统和动力学系统的吸引子、能量函数等模型和概念来描述和理解神经网络的结构和学习过程。当年夏天他应邀参加了位于巴黎的凡尔赛

生命研究所举办的主题为“从物理学到生物学”的研讨会,会后在得知主办方放弃了将会议论文结集出版的计划后,他便着手写作一篇适合在期刊发表的研究论文。霍普菲尔德心目中的读者群是物理学家和神经生物学家,《美国科学院院刊》是当时能兼顾二者的刊物^①。身为美国科学院院士,那时拥有的一项特权是每年可以不经同行审议即可在该刊发表一篇论文。此文于1982年元月中旬投稿,并于当年4月刊发。在此之后,霍普菲尔德与合作者接连在综合性的顶级期刊如《自然》《科学》和《美国科学院院刊》以及当时一直支持神经网络研究的专业期刊《生物控制论》等发表了多篇有很大影响的论文,对之前的模型进行了充实和扩展,特别是将之应用于处理联想记忆之外的认知任务,例如解决多重约束条件下的优化问题,这不啻于在神经网络研究领域刮起了一股强劲的旋风。^{[13]–[18]}

霍普菲尔德的系列研究之所以会成为推动1980年代神经网络研究大复兴的扛鼎之作,得益于一系列“天时地利人和”的有利因素。首先,经过了十余年的寒冬期,神经网络研究已经于1970年代末走向了复苏。芬兰科学家托伊沃·科霍宁(Teuvo Kohonen)于1977年出版了《联想记忆:系统-理论进路》,这是运用自组织神经网络探究联想记忆的重要工作;^[19]1979年辛顿和安德森组织的“联想记忆的并行模型工作坊”在加州的圣迭戈召开,参会论文于1981年结集出版;^[11]甘利俊一和迈克尔·阿比卜(Michael Arbib)于1982年2月在东京组织了一场日美双边研讨会,会议论文集以《神经网络中的竞争与合作》为名在当年出版;^[20]杰罗姆·费德曼(Jerome Feldman)和达纳·巴拉德(Dana Ballard)的长文“联结主义模型及其性质”于1982年在《认知科学》期刊发表,使得“联结主义”一词得以大行其道,与神经网络研究的复兴交相呼应。^[21]

其次,霍普菲尔德在1980年之所以离开普林斯顿大学转投加州理工学院,原因之一是前

①在2024年12月8日所作的诺贝尔讲座中,霍普菲尔德说他当时心目中的读者群还包括计算机科学家。

者当时的算力资源和计算机支撑环境相当有限和局促,而后者则优越得多,从而使他可以方便地开展计算模拟研究,这对于他能够迅速完成1982年论文的工作至关重要;从普林斯顿大学物理系转到加州理工学院具有重视与物理学的交叉融合传统的生命科学部,也令霍普菲尔德倍感轻快,有关学科和专业身份的种种顾虑得以一扫而光。而那时的加州南部地区,正在成为神经网络研究的聚集地;除了加州理工学院,加州大学圣迭戈分校一度成为计算神经科学和联结主义的大本营,而神经网络研究的重要领军人物之一的阿比卜也于不久后加盟了位于洛杉矶的南加州大学。

除了天时和地利,“人和”因素对于霍普菲尔德大获成功同样助力非凡。就在其1982年论文发表之后不久,加州理工学院与贝尔实验室联合举办了名为Hopfest的专题会议,围绕此文展开深入研讨^①。这个邀请制的小型会议吸引了不同学科越来越多研究者的关注,在连续举办了几年之后,于1987年正式升级为“神经信息加工系统”年度会议(曾简称为NIPS,2018年更改为NeurIPS),如今是机器学习、人工智能和计算神经科学等领域最高级别和最具影响力的会议之一。此外,前文已提到,霍普

菲尔德在加州理工学院将他与另外两位重量级教授合作开设的“计算之物理学”课程全面升级为“计算与神经系统”的研究生专业,并由他担任负责人。这是第一个计算神经科学的博士专业点,其成功运行为神经网络和计算神经科学等领域源源不断地输送高级专业人才。对霍普菲尔德的跨学科研究最重要的支持力量来自物理学家群体,尤其是在其论文发表后的头几年。图1是根据Web of Science数据库所收录的引用了该文的文献,按照其学科分类导出的最初五年的数据分布情况,排名第一的是物理学,以微弱优势领先于计算机科学,而如果将图中所列类别中的光学并入到物理学,则这一优势可以进一步扩大;图2是最初十年的引用文献的学科分布情况,排名首位的是计算机科学(约为1000篇),物理学(约为500篇)以微弱优势领先于神经科学排名第二,而如果加上光学(约250篇),则可以呈现出明显优势。

由物理学进入生命科学领域并取得非凡成就,开先例者并非霍普菲尔德。马克斯·德尔布吕克(Max Delbrück)受到尼尔斯·玻尔(Niels Bohr)的启发和鼓励,于1930年代中期从物理学转向生物学,因对于噬菌体的研究获得1969年诺贝尔生理学或医学奖,他生前曾在

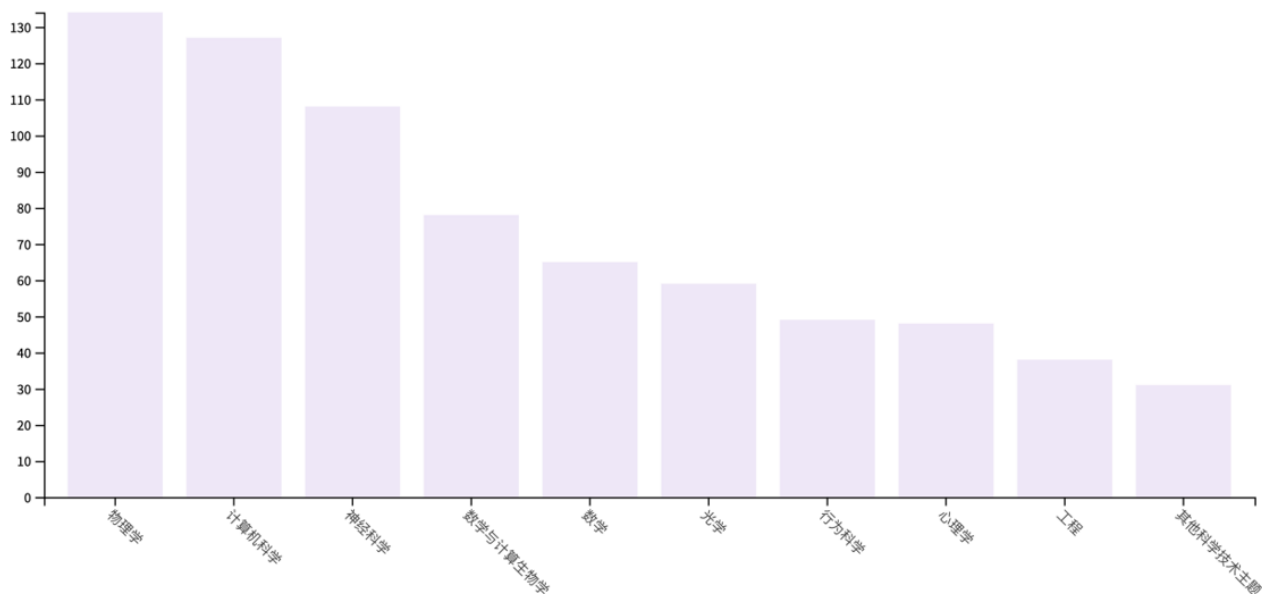


图1 霍普菲尔德1982论文发表五年后其引用数据的学科分布

^①Hopfest的本义是“啤酒节”或“啤酒会”,此处的用意很显然是与Hopfield之名相呼应,是对霍普菲尔德的致敬。

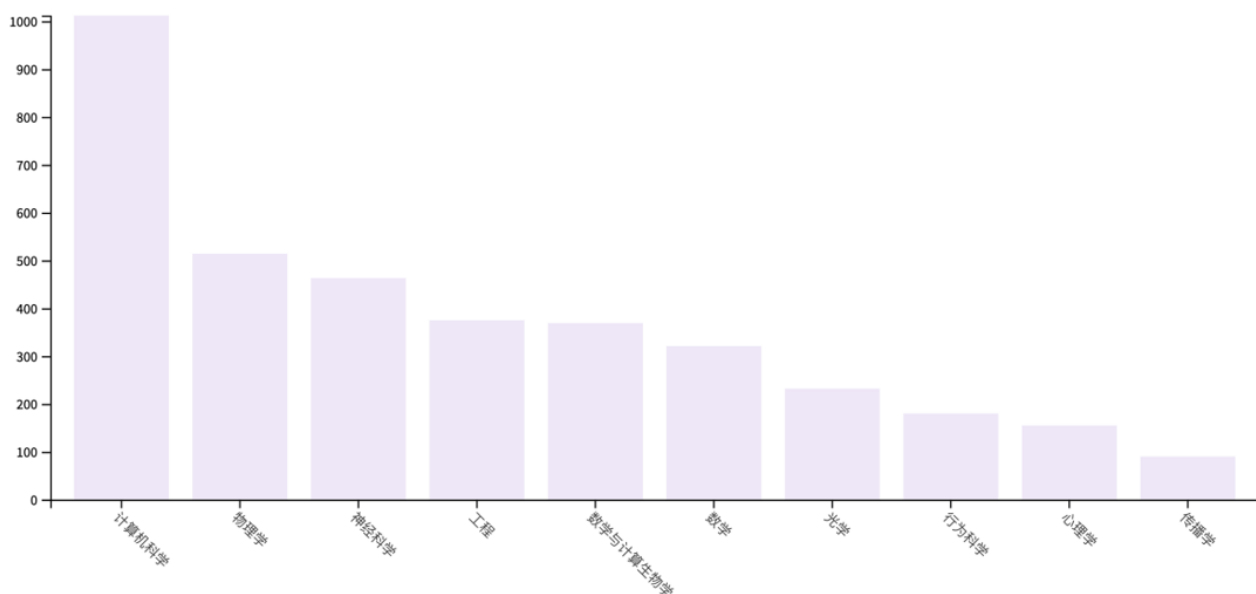


图2 霍普菲尔德1982论文发表十年后其引用数据的学科分布

加州理工学院任教30年。受其影响，物理学家薛定谔写出了《生命是什么？》这部跨界的传世之作。1962年因发现DNA分子结构而与詹姆斯·沃森（James Watson）分享诺贝尔生理学或医学奖的弗朗西斯·克里克（Francis Crick）和莫里斯·威尔金斯（Maurice Wilkins）也是物理学家出身。因提出被称为BCS（Bardeen-Cooper-Schrieffer）理论的经典超导学说，利昂·库珀（Leon Cooper）与其合作者分享了1972年的诺贝尔物理学奖；次年他在布朗大学创建了脑与神经系统研究所，从物理学转向神经科学，在脑与神经科学的研究上也做出了重要贡献，例如在1982年与其指导的博士生合作提出了关于神经元联结可塑性的BCM（Bienenstock-Cooper-Munro）理论。2024年的“大脑奖”（The Brain Prize）授予了拉里·艾伯特（Larry Abbott）、谢诺夫斯基和海姆·桑波林斯基（Haim Sompolinsky），以表彰他们在计算和理论神经科学领域做出的开创性贡献；这是神经科学学科最高级别的国际性大奖，而本次获奖的三人均为物理学家出身，在获得博士学位甚至在多年从事物理学研究之后，才转到了神经科学。对于霍普菲尔德跨界转行研究计算神经科学，物理学家不仅乐观其成，还有不少人积极跟进。受到霍普菲尔德的影响而转入神经网络研究的物理学家，不仅有不少当时的博士后和青年教

师，也有一些已成名的中年学者。例如丹尼尔·阿米特（Daniel Amit）、哈诺克·古特弗伦德（Hanoch Gutfreund）和桑波林斯基“三人组”依托统计物理学对霍普菲尔德网络的结构和算法进行了拓展，在1980年代中期曾联名发表了多篇有影响的论文，前二者均为1930年代出生；此三人后来均成为知名的计算神经科学家。

在获得2024年诺贝尔物理学奖之前，霍普菲尔德已荣膺多项科学大奖，其中多数为物理学的重大奖项，并曾于2006年担任过美国物理学会主席，这表明物理学是一门胸襟开阔的学科，对于跨学科研究具有很强的包容性。发表于2015年的一篇题为“一个世纪的物理学”的论文，以Web of Science所收录的百余年的四千万篇文献为基础进行数据挖掘和分析，根据每篇论文的参考文献和被引用模式来确定其是否为物理学论文，结果显示在242种物理学专业期刊所发表的论文总数为240万篇，而在上述物理学核心期刊之外的期刊所发表的物理学论文为320万篇，这表明发表在非物理学期刊（含Nature和Science这样的综合类期刊）上的物理学论文，实际上比发表在物理学专业期刊上的数量更多；^[22]基于进一步的量化分析，该文还表明近几十年物理学的发展趋势在总体上是变得更加多学科化和跨学科化，有6篇获得诺贝尔物理学奖的论文可以归入跨学科类

别,并提出霍普菲尔的1982年论文是有望获得诺贝尔物理学奖的热门候选者之一。2021年诺贝尔物理学奖颁发给三位对“理解复杂物理系统所做出的开创性贡献”的科学家,而其中两位真锅淑郎(Syukuro Manabe)和克劳斯·哈塞尔曼(Klaus Hasselmann)的身份是物理学出身的气象学家和海洋学家,连同2024年的“惊世骇俗”之选,似乎也可以反映出当代物理学逐渐增强的跨学科趋向。

那么,将2024年物理学奖单独授予霍普菲尔一人,是否也合情合理呢?他于1982年发表的论文催生了神经网络研究在1980年代的复兴,开启了一个新的时代,也因此多次获得物理学的其他大奖,再增加一顶诺奖桂冠,似乎也顺理成章,属于锦上添花。然而,就此独享尊荣,也并非全无顾虑和质疑。首先,提出霍普菲尔网络固然具有重大的历史意义,而且至今因其优雅简明的结构和精致的数学描述而持续吸引物理学家和计算机科学家的钻研热情,但这个模型几乎毫无工程上的实际应用价值,与近年来大显神威的机器学习相去甚远。其次,1980年代神经网络研究的复兴在十余年后再次跌入低谷,在相当长的时间里乏人问津。在2010年前后深度学习革命发生之前的十几年期间,机器学习领域最活跃的是以支持向量机(support vector machine)为代表的统计学习方法。由IBM团队开发的人工智能系统Watson,2011年在美国百科知识竞赛电视节目Jeopardy!中一举击败了两位人类顶级选手,引起轰动;当年的IBM Watson集成了彼时人工智能的众多传统技术,堪称当今人工智能的大数据、大模型、大算力和高能耗发展路线的先驱。倘若没有随后发生的深度学习革命,基于神经网络的机器学习恐怕至今仍偏安一隅,而霍普菲尔的历史性贡献固然不可磨灭,但还不至于因此而彪炳青史。

2. 玻尔兹曼机

2024年12月8日,在当年的诺贝尔奖颁授仪式之前,辛顿在瑞典斯德哥尔摩大学以“玻

尔兹曼机”为题作了诺贝尔讲座,尝试以浅显易懂的语言介绍他在神经网络研究上的这一重要工作,以突出统计物理学对于神经网络和机器学习所产生的关键影响^①。辛顿首先介绍了霍普菲尔网络,其简化版本由二值(状态为0或1)神经元组成,每个神经元与其他神经元之间均有对称的加权连接;霍普菲尔引入了能量函数等概念来描述网络的宏观行为,使用赫伯学习算法对其连接权重进行重复更新迭代,网络最终会稳定到某个能量最小值;霍普菲尔网络可用于模拟联想记忆等认知活动。玻尔兹曼机最初的提出可视为对霍普菲尔网络的升级:网络结构更为复杂,神经元区分为可见的与隐藏的两类;神经元状态可分为“睡眠”和“清醒”两种形态;引入随机性因素来增加网络稳定于全局最小能量值的概率;学习算法更加丰富和精致;网络具有一定的生成能力,即能够产生出新的内容;从统计物理学引入热平衡等概念来理解网络的宏观行为,系统稳定后并非处于某个状态,而是符合特定的概率分布,称为玻尔兹曼分布。玻尔兹曼机的局限性是计算效率不高,训练过程缓慢。为提升学习效率,加速网络达到热平衡的进程,辛顿等人提出了受限的玻尔兹曼机,对玻尔兹曼机进行了大量限制,例如只保留彼此之间没有连接的隐藏单元,从而得到收敛速度更快的学习算法。尽管其应用范围有限,受限的玻尔兹曼机仍取得了一定的实际应用,例如在为奈飞公司开发的推荐系统以及一些语言识别软件中都有过技术上的突破,但后续发展乏力,最终陷于停滞。受限的玻尔兹曼机的历史意义,用辛顿的话来说,是成为催生了深度学习革命的“酶”,让人们认识到深度神经网络是可行的,前途无限;而一旦完成了这个转变,这种“酶”也就不再需要了。

玻尔兹曼机是辛顿特别钟情的研究项目,也是令他深感自豪的重要成就,从1983年发表第一篇论文开始,前后共计发表了约40篇研究论文,时间跨度超过了30年。早期的主要合作

^①参见相关网址:<https://www.nobelprize.org/prizes/physics/2024/hinton/lecture/>。

者是物理学出身的谢诺夫斯基，他们二人合作发表过不少于5篇论文；每每谈及玻尔兹曼机，辛顿总是会提到谢诺夫斯基，强调他的贡献，例如在他30分钟的诺贝尔演讲中，共有4次提到了谢诺夫斯基。从霍普菲尔德网络到玻尔兹曼机，二者之间的桥梁不仅有来自统计物理学的理论和概念，还有谢诺夫斯基这位关键人物，他是霍普菲尔德指导的物理学博士，更是辛顿的终身挚友和学术同道，作为计算神经科学的领军人物，他在人工智能和认知科学领域做出过重要贡献。

3. 深度学习革命

发生于2010年代初期的深度学习革命，是人工智能发展史上的最为重大的突破之一，也是本世纪迄今为止最具影响力的科技革命。关于这场革命的历史叙事和发展脉络，重要的思想来源、技术突破和发现/发明的首创权的辨析和认定，以及关键人物的历史性贡献的评价和荣誉归属，都存在不同的叙述和解读，有些甚至成为聚讼纷纭、争论不休的公案。^[23]而辛顿无疑是深度学习革命最具代表性的人物。首先，辛顿从1970年代中期投身神经网络研究，1979年身为博士后曾作为新生代的骨干，参与推动关于联想记忆研究的神经网络进路的复苏，成为1980年代神经网络研究复兴浪潮中引人注目的新锐力量，做出了多项重要工作。其次，在1990年代中后期神经网络研究再度陷入低潮期时，辛顿初衷不改，依托加拿大高等研究院提供的宝贵支持，聚集了杨立昆和本吉奥等青年英才，在较为困难的情况下，经过十余年的勉力坚守，最终于2006年依托受限的玻尔兹曼机推出了深度信念网络，为随后兴起的深度学习革命指引了方向。^[13]第三，得益于英伟达公司开发的图形处理器（GPU）被应用于训练神经网络，辛顿与两位技术能力出众的博士研究生合作开发了名为AlexNet的大型卷积神经网络系统，于2012年在基于图像数据库ImageNet的图像识别挑战赛中一举夺魁，并且领先第二名10个百分点。这一突破性的成就震惊了当时的机器学习和人工智能领域，由此揭开了深度学习革命的大幕。

选择辛顿与霍普菲尔德共享2024年诺贝尔物理学奖，堪称珠联璧合的完美组合。霍普菲尔德的物理学家本色及其运用物理学理论资源解决计算神经科学问题所展现的物理学思维和技能，有助于保持物理学奖的底色和尊严，并展示了物理学对于当代人工智能发展所做出的历史性贡献。辛顿虽然不是物理学家，但他的贡献与霍普菲尔德的工作却并非脱节，而是具有思想、方法和社会关系上的关联性和连续性，而他40多年贯穿始终近乎打满全场的独特经历，以及多次在重要历史关头所做出的卓越抉择和突破性成就，让他这位“深度学习之父”和“当代人工智能教父”加冕诺贝尔物理学奖的桂冠，似乎也顺理成章，当之无愧。

三、剽窃疑云、首创权之争与荣誉归属

2024年物理学奖所引发的另一争议是关于学术不端与失范、重要发现和发明的首创权之争以及科学荣誉归属和分配等一系列学界内部问题，牵涉的人物众多，涉及的问题也较为复杂，有待开展全面深入的调查研究和周密分析。限于篇幅，本文只能触及其中的部分问题，并尝试解析一二。

1. 剽窃疑云与首创权之争

剽窃疑云所涉者主要为霍普菲尔德与日本科学家甘利俊一。在霍普菲尔德的1982年文章发表数年之后，陆续有研究者发现甘利俊一早在1972年发表的论文中就已经提出了同样的网络，并用于模拟联想记忆。^[24]从1990年代中期开始，文献中陆续出现了“Amari-Hopfield Network”这样的称谓。吴思在题为“2024诺贝尔物理学奖是否出现了重大遗漏？”的文章中写道：两篇论文中的数学模型几乎一模一样，包括神经元的阈值动力学和神经元连接的学习律，而且甘利俊一的文章还做了更为细致的数学分析；基于该数学模型，两篇文章都分析了网络动力学的稳定状态，即吸引子，并由此引申到了大脑的联想式记忆；仔细对比两篇论文，“有科研经验的读者都知道，在模型如此相似的情况下，有了甘利俊一1972年的文章，霍普

菲尔德1982年的文章其实很难能发表在重要杂志上了。当然在当时资讯条件下,有可能霍普菲尔德教授并不知道甘利俊一教授的工作。”^[25] 吴思指出:甘利俊一比霍普菲尔德早十年提出该模型,在计算神经科学及相关领域,并不是一个鲜为人知的秘密,而是很多人都知道的事实^①。

甘利俊一出生于1936年,1963年获东京大学工学部数理工程博士学位。在九州大学任职数年之后,他于1967年担任东京大学工学部教职,直至1996年退休,成为名誉教授。此外,甘利俊一还长期任职于日本国立理化学研究所。自1960年代中期起,甘利俊一就致力于神经网络研究,这在当时正值长达十几年的寒冬期,殊为不易;他的多项研究成果是先驱性的,超前于时代,而他开创的信息几何学具有广泛的影响。他曾经获得国际电气电子工程师学会颁发的神经网络先驱奖和伊曼纽尔·皮欧雷奖(IEEE Emanuel R. Piore Award),并于2024年入选日本学士院。

在其1982年的论文中,霍普菲尔德并未引用甘利俊一的1972年论文,而是引用了后者此后发表的两篇论文。在2024年诺贝尔物理学奖颁发后,甘利俊一在接受访谈时提及此事:“事实上,如果看了霍普菲尔德的论文,就会发现其中引用了我的论文,但被引用的是没有直接关联的论文。我的联想记忆论文并未被直接引用,引用的是我后来写的关于自组织的论文。我不知道为何会出现这样的引用错误,不过这种事情确实会发生。”^[26]

霍普菲尔德当时真的是发生了参考文献引用上的错误吗?问题的关键是那时的霍普菲尔德是否知悉并阅读过甘利俊一的1972年论文。根据谷歌学术的统计,此文发表后的头十年只有15次引用,其中半数还是自引,而那时的霍普菲尔德转入神经网络和联想记忆的研究时间不长,很有可能没有注意到此文。尽管在霍普菲尔德论文里所引用的甘利俊一的两篇论文之

一,通过参考文献可以“顺藤摸瓜”查找到此文,但此文的标题却并不容易让人辨认出这是与联想记忆有关的工作。而霍普菲尔德的论文未经同行专家评审就得以发表,也失去了在审稿过程中得到提醒或被要求补上引用甘利俊一论文的机会。

1970年代是人工智能发展的低潮期,而对于神经网络研究而言,其所处的学术环境则更为严峻。甘利俊一在极为不利的形势下,持续不断坚持研究,并做出了多项超前其时代的重要成果,着实难能可贵,但时代的困局终究难以破解,这导致甘利俊一在当时所做出的若干项先驱性研究成果,尤其是其1972年发表的论文,未被多数同行注意到,可谓生不逢时。福岛邦彦(Kunihiko Fukushima)是甘利俊一的同时代人,其学术经历也颇为相似,都是从1960年代中后期起开展神经网络研究,而他最重要的成果发表于1980年,时值神经网络研究复兴的预热期,其境遇与甘利俊一相比则大相径庭:据谷歌学术统计,该论文发表后的头十年获得了约150次引用,而到2025年6月,总引用数超过1万次。^[27] 福岛邦彦也因此项成果被公认为卷积神经网络的首创者和深度学习研究的先驱。

当时的霍普菲尔德很有可能并不知悉甘利俊一的1972年论文,但这足以表明霍普菲尔德是独立作出了极为相似的发现,因而可以免除剽窃之嫌吗?情况恐并非这么简单。在霍普菲尔德1982年论文所引用的甘利俊一的两篇论文当中,有一篇值得一提。这篇发表于1977年的论文的标题是“联想与概念形成的神经理论”,其主要内容是基于甘利俊一此前的系列研究成果,尝试提出关于联想与概念生成的具有整合性的神经计算机制。^[28] 虽然不少技术细节被省略,但甘利俊一1972年论文的要点还是被吸收进了这篇论文之中,因而可以说此文包含了后来被称为霍普菲尔德网络的结构和数学表达的

①吴思现为北京大学心理与认知科学学院教授,并担任该学院院长,其研究领域为计算神经科学和类脑计算。他在北京师范大学完成了从物理学本科到博士的训练,并曾在多所国际性研究机构和大学从事博士后研究或担任教职,其中包括日本理化学研究所,与甘利俊一合作发表过多篇论文。

大部分要素^①。霍普菲尔德因此而获得重要启发，甚至借鉴了其核心内容，似乎是有可能的。前文提到的甘利俊一（在2024年底）所言霍普菲尔德的1982年论文所引用的是没有直接关联的论文，当为不确之论。

霍普菲尔德在文中将甘利俊一的这两篇论文，连同三篇其他作者的论文，作为刻画目标问题和方法论背景的参考文献，一笔带过，未曾介绍和讨论其具体内容和观点。其中之一是斯坦福大学的物理学家威廉·利透（William Little）于1974年发表的论文，提出神经网络中持续存在的稳定状态可以类比于统计物理学中的埃辛自旋系统（Ising spin system）中保持规则排列的有序状态，且二者的动态演变也具有相似性，可用于说明短时记忆和长时记忆的一些特性。^[29]利透的这篇论文可以说是更早将神经网络与埃辛自旋系统联系起来的工作，这对于为霍普菲尔德网络给出物理学解释很有启发。在某种意义上，霍普菲尔德1982年论文的工作可以说是对甘利俊一和利透先前成果的综合、拓展和提升，而霍普菲尔德的独特贡献是提出了能量函数的概念用于刻画网络的宏观性质。霍普菲尔德的论文获得了巨大的成功，其对前人的相关工作颇有机巧的引用处理，一方面算是对前人的贡献给予了一定的承认，尽管轻描淡写，未能尽显君子之风，但在学术规范的意义上似无明显瑕疵，或可有利于规避学术不端的嫌疑，另一方面则以不着痕迹的方式，突出或者放大了自己的贡献，最终造成了独享殊荣的结果。

甘利俊一和利透的首创权也并非完全没有得到同行的认可：在谷歌学术网站以“Little-Hopfield”/“Hopfield-Little”和“Amari-Hopfield”/“Hopfield-Amari”为关键词进行检索，截至2025年6月初，可分别得到约330篇（最

早出现于1985年）和220篇文献（最早出现于1995年）。甘利俊一在2013年发表的文章中提到，桑波林斯基在某篇论文中曾有过这样的评论：甘利俊一更早提出了同样的模型，但由于霍普菲尔德的贡献实在太太，所以我们称之为霍普菲尔德模型。^[30]据说阿米特生前曾当面给甘利俊一道歉，表示这是学术界的一个错误，该模型应当被称为Amari-Hopfield模型，并将一条红色领带赠与甘利俊一表达歉意。^[25]如前所述，物理学家出身的阿米特和桑波林斯基曾是1980年代中后期霍普菲尔德网络研究的重要推手，阿米特还在1989年出版了一部影响很大的著作《大脑功能建模：吸引子神经网络的世界》，书中仅有一次提及甘利俊一（引用其1972年论文），而霍普菲尔德则反复被提到，是书中出现次数最多的人名。^[31]可以设想一下，假若霍普菲尔德在其1982年的论文中，花费一定笔墨分别介绍甘利俊一和利透先前的工作，并给予较为充分的承认，而不只是作为背景材料一带而过，霍普菲尔德网络或有可能被命名为Amari-Little-Hopfield网络^②。

2. 首创权、科研奖励与荣誉归属

1957年8月，罗伯特·默顿（Robert K. Merton）就任美国社会学学会主席，于当年的学会年会上发表了题为“科学发现的首创权：科学社会学的一个篇章”的就职演讲，文稿随后发表在学会会刊《美国社会学评论》上，由此开启了学界关于首创权、科研奖励系统和荣誉归属机制等问题的研究，涉及科学史、科学社会学、科学哲学和科学计量学等多个学科。^[32]正如默顿所指出的，科学界对于首创权的计较、争夺以及由此引发的攻讦，远较其他行业为甚，在历史上和现实中都层出不穷，延绵不绝。许多科学界的代表性人物，包括伽利略、牛顿、莱布尼茨和爱因斯坦等在内，都曾主动或被动

①笔者曾将甘利俊一的1977年论文发给厦门大学物理科学与技术学院的赵鸿教授，请他判别是否可以说该文已包含了霍普菲尔德网络的要素，他回答说“是”，甘利俊一提出的网络可以说是霍普菲尔德网络的前身，但霍普菲尔德在此基础上也有重要创新。赵鸿教授的专长研究领域是非平衡统计物理学、非线性动力学和机器学习。

②最近确实已出现了如此冠名的文献：Irie K., Gershman J. ‘Fast Weight Programming and Linear Transformers: From Machine Learning to Neurobiology’ [EB/OL]. arXiv:2508.08435v1, <https://doi.org/10.48550/arXiv.2508.08435>, 2025-8。作者为哈佛大学的两位心理学家。

地卷入过关于首创权的纷争中;即便谦逊宽厚者如达尔文,也很在意首创权的归属。知识、方法成果的共享和公有,是科学共同体组织和运行的基本规范之一;原创性是科研活动所追求的关键指标,而对首创权的认定是对科学家创造性成就的肯定和承认,这种群体层面的评价性标准很自然地内化为科学家个体的职业成就动机,激励科学家投入时间和精力开展创新性的研究工作。对他人成果进行恰当的引用和介绍,对重要的发现、模型和方法等进行冠名,包括诺贝尔奖在内的各种级别和领域的科研奖励系统等,都是对原创性贡献给予认定和承认的通行方式。科学界行之有效的首创权原则、奖励系统和荣誉分配机制,有利于激励和保护创新,并促进科研资源(包括人力资源)的合理分配以及科学探索的多样性发展。^[33]然而,这种对首创权的强调也可能导致一些冷酷或消极的后果,例如“胜者为王”:只有“冠军”才能独享首创权,而其他做出过重要贡献的科学家(团队)所获得的承认和荣誉往往不成比例;验证性的和重复(再现)性的研究工作由于不被鼓励,间接导致(部分)科研成果的可重复性降低。^[34]

理论上,根据成果公开发表在时间上的先后似乎就能清楚明白地认定首创权,但在现实中,确认首创权的归属有时却并非易事,牵涉很多复杂因素。首先,有如不少研究者所发现的,多人分别独立做出同样的发现或发明的情况并非稀有,而是屡见不鲜。百年前的一项研究列举了近150项在科学技术上由多人分别独立做出同样创新的案例,除了牛顿与莱布尼茨各自独立发明微积分、达尔文与华莱士几乎同时提出自然选择理论等众所周知的历史往事,还包括重复发明摆钟(至少3次)、温度计(7次)、电报(4次)和雷达(6次)等情况。^[35]默顿进而指出,多人独立做出相同的发现或发明,在科技史上并非例外情况,而是普遍现象,其中有相当一部分是相隔多年以后的再次发现或发明。^[36]在当代神经网络和机器学习

领域被广泛运用的反传算法(backpropagation algorithm)曾被多人多次独立提出和发展,自1960年代初开始,先后有10余人被认定为反传算法的首创者或先驱者,其中包括甘利俊一,直到1986年经由罗姆哈特、辛顿和谢诺夫斯基等人的又一次发现和开发,反传算法的强大威力和适用性才被普遍了解和认可^①。

其次,正如默顿所揭示的,科学界在资源、荣誉和承认的分配和认定中较为普遍地存在所谓“优势累积效应”:已经具有成就、地位和名望的科学家往往会因此获得更多与之实际贡献和能力不相符的名誉、认可和资源,而相对处于弱势和较低地位的研究人员则被不当地忽视,导致他们本应获得的荣誉和认可被剥夺。^[37],^[38]默顿将之称为“马太效应”,取自《圣经·新约·马太福音》中的一个寓言:本就所有不多者,连他已有的也要剥夺;本已拥有很多者,还要给他更多,让他多多益善。社会学的实证研究表明马太效应不仅在科学界具有相当的普遍性,在其他很多行业和领域也不同程度的存在。^{[39]-[41]}如果两位科学家各自独立地做出了同一个重要发现或发明,其中较为知名者会获得更多的认可和荣誉,乃至独享首创权。这一明显具有不平等、不公正色彩的现象之所以在科学界持续存在,确有其深厚的社会认知基础。相较于籍籍无名者,已成名科学家的名字更容易吸引到同行的注意,也更容易被记住,更容易在群体中传播,从而如同滚雪球一般变得越来越壮大。不仅如此,虽然两位科学家独立取得了同样的进展或突破,但已成名科学家因其先前的成功所赢得的荣誉和地位,会令更多的同行相信其选题的眼光和洞察力,从而吸引到更多的跟随者。^[42]

第三,同样源于默顿的睿见,是关于科学界特有的冠名现象:特定学科或研究领域的开创者往往被称为“XX之父”,而各种定律、定理、假说、常数、方程、不等式以及实验装置等,常以首创者或发明者的名字来命名。^[32]然而,统计学家斯蒂芬·斯泰格勒(Stephen Stigler)

①参见关于反传算法的介绍:<https://en.wikipedia.org/wiki/Backpropagation>.

在1980年的一篇论文中提出了“斯泰格勒冠名定律”，声称“没有一项科学发现是以原初的发现者来冠名的”。^[43]这种张冠李戴的情况在历史上并不罕见，例如众所周知的毕达哥拉斯定理并非毕达哥拉斯最先发现或证明的，而且毕达哥拉斯本人的数学家身份也极为可疑。^[44]在斯泰格勒所熟知的数理统计学领域，类似情况也不胜枚举：拉普拉斯在傅里叶发表其研究成果之前，已在其论著中运用了后来被称为傅里叶变换的方法；拉格朗日早在拉普拉斯开始其科学生涯之前，就已经给出了（后来才被命名的）拉普拉斯变换；在柯西不经意间触及相关问题的近30年前，泊松早在1824年就已经发布了被后世命名的柯西分布；凡此种种，不一而足。斯泰格勒不无诙谐地指出，斯泰格勒定律本身也与其内容是相符的：默顿在其1957年的就职演讲中已经表达了相似的观点，而以“斯泰格勒”而不是“默顿”来冠名，就是为了让这条定律也能够名实相符。斯泰格勒定律的表述固然有夸大其词之嫌，但也确实揭示了科学界冠名现象的部分实情：冠名的提出和传播过程并非总是遵循严谨公正、实事求是的标准，而是具有一定的随意性和偶然性，掺杂了诸多复杂社会因素的影响。^{[45]–[48]}由于冠名意味着对首创权的认定以及相应的荣誉和知名度，对后续的科学奖励和荣誉分配自然有可能产生直接影响。对科学界冠名现象及其效应的反思和评价，近期也引发了关于其存废和改良问题的争议。^{[49]–[51]}

基于以上对由默顿科学社会学理论引发的几个论题的叙述和分析，不难看到2024年诺贝尔物理学奖的两位得主，尤其是霍普菲尔德，在通往诺奖之路上所能借助的种种有利的社会因素：30岁出头即成为普林斯顿大学的物理学正教授，40岁入选美国国家科学院，早期职业生涯的顺风顺水，令他有底气和信心从容不迫地尝试跨学科转型，找寻能适配他才能和雄心的重大科学问题；而一旦获得突破，便积极利用先前具备的累积优势，充分调集各种资源，迅速扩大战果，形成一呼百应之势，将冠名权、首创权等承认和荣誉悉数收入囊中，各种重要

奖项随后便纷至沓来，最终水到渠成得以问鼎诺奖。

“好风凭借力，送我上青云。”任何一项具有重大创新性的科研成果得到同行的关注和承认，都需要生逢其时，具备天时地利人和的有利条件。科学计量学家安东尼·范瑞安（Anthony van Raan）在2004年的一篇论文中提出了科学中的“睡美人现象”：有些具有创新性的科研成果发表后，在相当长的一段时间内几乎无人问津，默默无闻，成为沉睡中的美人，直到某一天被王子唤醒，忽然间获得众人的注目。^[52]科学中的睡美人现象最具代表性的案例之一是孟德尔于1865年发表的植物遗传学研究，这项超前于时代的先驱性成果直到30多年后才引起了生物学家的注意。按照范瑞安提出的睡美人的定义标准，对当时收录了2000万篇论文的数据库进行统计分析，得出的结论是在每年发表的科技文献中，睡美人论文的比例是0.01%。而10年后的一项更为系统性和精细化的研究则表明，睡美人论文的比例比范瑞安的估算值要高出100倍以上，且在不同学科的分布有差异。^[53]科学计量学关于睡美人现象的研究所统计的仅是被“唤醒”的睡美人论文，而如果考虑到还有不少沉睡中的美人未能有幸被王子唤醒，睡美人论文所占的比例还有可能更高。演化生物学家安德烈亚斯·瓦格纳（Andreas Wagner）将“睡美人”概念用于解释自然界的生物演化和人类的文化演化（包含科学与技术），以阐明在物种演化过程中依托基因突变等因素所产生的新物种，往往需要经历漫长的蛰伏期，等到与其所具备的竞争优势相适应的时代到来，才迎来蓬勃发展的新局面，而这种情况在人类的文化、艺术、技术和科学的创新活动及其发展过程中，也同样屡见不鲜。^[54]例如草类植物是当今最为成功的植物种类之一，遍布地球的大多数地方，有多达一万多个品种，其起源早在6500多万年前恐龙时代，但经历了漫长的艰难求生的困境，直到2500万年前才成为如今这般的主导型物种。这样的案例在生物演化的历程中不胜枚举，表明成功的物种演化不仅取决于内在的“创新”所提供的适用性或竞争性

优势,也依赖其所处的环境和世界。人类的创新活动也是如此,许多超前于时代的创新会遭遇“睡美人”的境遇,只有等到与之相适应的思想观念和认知、社会体制和氛围的逐渐成型,才有可能展现其威力和优势。

与自然界演变的无目的性有所不同,人类可以通过运用自身的能动性有目的地参与营建有利于创新及其接纳的认知与社会环境。前文已介绍和分析过霍普菲尔德在其1982年论文发表后,及时跟进开展的一系列“造势”活动。2006年,在依托受限的玻尔兹曼机提出深度学习的思想之后,辛顿和杨立昆等人申请在次年的NeurIPS会议上举办一次专题研讨会,但被会议组织者在没有做任何解释的情况下否决了。辛顿等人于是利用加拿大高等研究院提供的经费组织了“私设”的研讨会,并租用大巴来接送对深度学习有兴趣的NeurIPS参会者,成功地吸引了300人前往,这是“深度学习”这一术语在业内的首次公开亮相。^[55]大约在2010年,深度学习在语音识别领域实现了突破,取得了第一批成果。辛顿提出了一个绝妙的想法:在接下来的夏季实习期内,将三名博士生分别送到了业内最领先的三家大公司——谷歌、微软和IBM,并让他们采用深度神经网络替换原有的系统;果不其然,系统性能有了显著提升。18个月后,这三家公司都部署了新的基于深度学习的语音识别系统。辛顿等人精准施策的“造势”活动为推动深度学习革命的到来起到了至关重要的作用。

结 语

2024年诺贝尔物理学奖的颁授,其人选和获奖成就引发了广泛的关注、疑惑和争议,其原因在于公众受制于基于普及性教育所形成的关于物理学的刻板印象,对于物理学对当代人工智能的发展所做的贡献几无所知,且对于物理学所具有的跨学科特性以及学科边界的模糊性和动态演变,也缺乏必要的了解。诺贝尔奖委员会所做出的出乎意料的决定,或可有助于弥补和纠正先前的忽视以及公众并不精准的刻

板印象。

与2024年诺贝尔物理学奖有关的另一项争议涉及科学发现或发明的首创权的认定,及其与科学奖励和荣誉分配之间的关联。本文对此进行了深入细致的分析,并提出首创权并非科研奖励和荣誉分配中的唯一标准,而促成科技创新得以完成、验证、实现和接受的重要贡献,也同样值得重视。2024年诺贝尔物理学奖的两位得主,尽管未必能分别独享相关创新的首创权,甚至有关剽窃疑云的阴影犹在,但却无法否认他们所做出的远胜他人的重要贡献和所产生的重大历史影响,由此观之,他们最终得以蟾宫折桂,应属当之无愧。

对于2024年诺贝尔物理学奖已有数量颇为可观的解读文献问世,但关注其背后难以觉察到的复杂社会因素的研究却难得一见,本文弥补了这方面的不足,侧重于对与之有关的复杂社会因素进行细致的探究,揭示了先前的诸多解读性文章所忽视的关键细节和事件关联,展现了更为清晰的历史脉络和人物风貌。这对于更好地理解当代科学的社会运行,当不无裨益。

尽管本文重点探讨的是科研首创权认定与科学荣誉分配等科学社会学的传统议题,以及公众理解科学等科学技术与社会研究的相关问题,但总体上本文的研究应当属于关于当代科学技术史的一个具体案例解析。当代科技史研究往往不受史家青睐,主要原因在于研究对象未经历史的洗涤和沉淀而缺少必要的观察距离和稳定性,相关资料繁杂浩瀚、良莠不齐,以及所涉人物或机构的利益相关性等。但开展当代科技史研究也有相应的优势:资料丰富,大多数当事人和见证者尚在人世,有条件获取有价值的反馈以澄清争端或将探究引向深入,可以及时抢救和收集关键证据,等等。^{[56], [57]}本文的研究未开展针对关键当事人的问询和访谈,因而未能获取并提供相关的反馈和更有针对性的信息,但也因此保留了较大程度的中立性和必要的距离。秉持“史家乃侦探”的研究策略,借助当今互联网和融媒体时代极为丰富多样的数字化资料,本文尝试在浩瀚庞杂的素材中去芜存菁,提取富有价值的键信息,构

建合理的叙事和解读，成一家之言而已。

[参考文献]

- [1] Castelvecchi, D., Callaway, E., Kwon, D. 'AI Comes to the Nobels: Double Win Sparks Debate About Scientific Fields' [EB/OL]. *Nature*, <https://www.nature.com/articles/d41586-024-03310-8>. 2024-10-10.
- [2] Metz, C. 'How Does It Feel to Win a Nobel Prize?' [EB/OL]. *The New York Times*, <https://www.nytimes.com/2024/10/08/technology/nobel-prize-geoffrey-hinton-ai.html>. 2024-10-08.
- [3] Coulter, M. 'Google's Nobel Prize Winners Stir Debate Over AI Research' [EB/OL]. *Reuters*. <https://www.reuters.com/technology/artificial-intelligence/googles-nobel-prize-winners-stir-debate-over-ai-research-2024-10-10/>. 2024-10-10.
- [4] 弗兰克·维尔切克. 从物理到智能: 当诺奖得主遇上诺奖得主 [EB/OL], 胡风、梁丁当译, 墨子沙龙, <https://mozi.ustc.edu.cn/detail/1201>. 2024-10-30.
- [5] 酆全民. 人工智能研究的物理学途径 [J]. 物理, 2001, (2): 90-95.
- [6] Jiao, L., Song, X., You, C. et al. 'AI Meets Physics: A Comprehensive Survey' [J]. *Artificial Intelligence Review*, 2024, 57(9): 256.
- [7] Schmidhuber, J. 'How 3 Turing Awardees Republished Key Methods and Ideas Whose Creators They Failed to Credit' [EB/OL]. <https://people.idsia.ch/~juergen/ai-priority-disputes.html>. 2023-12-14.
- [8] Hopfield, J. 'Now What' [EB/OL]. <https://pni.princeton.edu/people/john-j-hopfield/now-what>. 2025-10-10.
- [9] Hopfield, J. 'Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity' [J]. *Proceedings of the National Academy of Sciences*, 1974, 71(10): 4135-4139.
- [10] Hinton, G., Geoffrey, E. 'Hinton' [A], Anderson, J., Rosenfeld, E. (Eds.) *Talking Nets: An Oral History of Neural Networks* [C], Cambridge, MA: The MIT Press, 1998, 361-385.
- [11] Hinton, G. E., Anderson, J. A. (Eds.) *Parallel Models of Associative Memory* [C], Hillsdale, NJ: Erlbaum, 1981.
- [12] Rumelhart, D. E., McClelland, J. L., PDP Research Group. *Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations* [M]. Cambridge, MA: The MIT Press, 1986.
- [13] Hinton, G., Osindero, S., Teh, Y. 'A Fast Learning Algorithm for Deep Belief Nets' [J]. *Neural Computation*, 2006, 18(7): 1527-1554.
- [14] Hopfield, J. 'Neural Networks and Physical Systems with Emergent Collective Computational Abilities' [J]. *Proceedings of the National Academy of Sciences*, 1982, 79(8): 2554-2558.
- [15] Hopfield, J., Feinstein, D., Palmer, R. 'Unlearning' Has a Stabilizing Effect in Collective Memories' [J]. *Nature*, 1983, 304(5922): 158-159.
- [16] Hopfield, J. 'Neurons with Graded Response Have Collective Computational Properties Like Those of Two-state Neurons' [J]. *Proceedings of the National Academy of Sciences*, 1984, 81(10): 3088-3092.
- [17] Hopfield, J., Tank, D. 'Neural' Computation of Decisions in Optimization Problems' [J]. *Biological Cybernetics*, 1985, 52(3): 141-152.
- [18] Hopfield, J., Tank, D. 'Computing with Neural Circuits: A Model' [J]. *Science*, 1986, 233(4764): 625-633.
- [19] Kohonen, T. *Associative Memory: A System-Theoretical Approach* [M]. Berlin: Springer-Verlag, 1977.
- [20] Amari, S., Arbib, M. (Eds.) *Competition and Cooperation in Neural Nets* [C], Berlin: Springer-Verlag, 1982.
- [21] Feldman, J., Ballard, D. 'Connectionist Models and Their Properties' [J]. *Cognitive Science*, 1982, 6(3): 205-254.
- [22] Sinatra, R., Deville, P., Szell, M., et al. 'A Century of Physics' [J]. *Nature Physics*, 2015, 11(10): 791-796.
- [23] 詹好. 审视深度学习革命: 关键人物的历史性贡献及其荣誉归属 [D]. 厦门: 厦门大学, 2025.
- [24] Amari, S. 'Learning Patterns and Pattern Sequences by Self-organizing Nets of Threshold Elements' [J]. *IEEE Transactions on Computers*, 1972, 21(11): 1197-1206.
- [25] 吴思. 2024诺贝尔物理奖是否出现了重大遗漏? [EB/OL], https://mp.weixin.qq.com/s/vkX-p3FTajp3WyG_LMoi8Q. 2024-12-03.
- [26] 宇津木聡史. トップを走っていた日本の神経回路網研究-AI研究のパイオニア、甘利俊一さん [前編] [EB/OL], *Science Portal*, https://scienceportal.jst.go.jp/explore/interview/20250131_e01/index.html. 2025-01-31.
- [27] Fukushima, K. 'Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position' [J]. *Biological Cybernetics*, 1980, 36(4): 193-202.
- [28] Amari, S. 'Neural Theory of Association and Concept-Formation' [J]. *Biological Cybernetics*, 1977, 26(3): 175-185.
- [29] Little, A. 'The Existence of Persistent States in the Brain' [J]. *Mathematical Biosciences*, 1974, 19(1-2): 101-120.

- [30] Amari, S. 'Dreaming of Mathematical Neuroscience for Half a Century' [J]. *Neural Networks*, 2013, 37: 48-51.
- [31] Amit, D. *Modeling Brain Function: The World of Attractor Neural Networks* [M]. Cambridge: Cambridge University Press, 1989.
- [32] Merton, R. K. 'Priorities in Scientific Discovery: A Chapter in the Sociology of Science' [J]. *American Sociological Review*, 1957, 22(6): 635-659.
- [33] Strevens, M. 'The Role of the Priority Rule in Science' [J]. *Journal of Philosophy*, 2003, 100(2): 55-79.
- [34] Romero, F. 'Novelty versus Replicability: Virtues and Vices in the Reward System of Science' [J]. *Philosophy of Science*, 2017, 84(5): 1031-1043.
- [35] Ogburn, W., Thomas, D. 'Are Inventions Inevitable? A Note on Social Evolution' [J]. *Political Science Quarterly*, 1922, 37(1): 83-98.
- [36] Merton, R. K. 'Singletons and Multiples in Scientific Discovery: A Chapter in the Sociology of Science' [J]. *Proceedings of the American Philosophical Society*, 1961, 105(5): 470-486.
- [37] Merton, R. K. 'Matthew Effect in Science: The Reward and Communication Systems of Science are Considered' [J]. *Science*, 1968, 159(3810): 56-63.
- [38] Merton, R. K. 'Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property' [J]. *Isis*, 1988, 79(4): 607-623.
- [39] Perc, M. 'The Matthew Effect in Empirical Data' [J]. *Journal of the Royal Society Interface*, 2014, 11(98): 20140378.
- [40] Katchanov, L., Markova, Y., Shmatko, A. 'Empirical Demonstration of the Matthew Effect in Scientific Research Careers' [J]. *Journal of Informetrics*, 2023, 17(4): 101465.
- [41] Rigney, D. *The Matthew Effect: How Advantage Begets Further Advantage* [M]. New York: Columbia University Press, 2010.
- [42] Strevens, M. 'The Role of the Matthew Effect in Science' [J]. *Studies in the History and Philosophy of Science (Part A)*, 2006, 37(2): 159-170.
- [43] Stigler, M. 'Stigler's Law of Eponymy' [J]. *Transactions of New York Academy of Sciences*, 1980, 39(1): 147-157.
- [44] Huffman, C. 'Pythagoras' [EB/OL]. Stanford Encyclopedia of Philosophy, <https://plato.stanford.edu/entries/pythagoras/>. 2024-02-05.
- [45] Strous, R., Edelman M. 'Eponyms and the Nazi Era: Time to Remember and Time for Change' [J]. *Israel Medical Association Journal*, 2007, 9(3): 207-214.
- [46] Kondziella, D. 'Thirty Neurological Eponyms Associated with the Nazi Era' [J]. *European Neurology*, 2009, 62(1): 56-64.
- [47] Beretta, M. 'Names as Rewards: The Ambiguous Role of Eponyms in the History of Science' [J]. *Nuncius*, 2019, 34(2): 219-235.
- [48] Poulin, R., McDougall, C., Presswell, B. 'What's in a Name? Taxonomic and Gender Biases in the Etymology of New Species Names' [J]. *Proceedings of the Royal Society B*, 2022, 289(1974): 20212708.
- [49] Schulman, F., Rissi, R. 'Eponyms in Science: A Proposal to Minimize Their Use' [J]. *Veterinary Pathology*, 2024, 61(5): 846-848.
- [50] Maxie, G. 'Response to "Eponyms in Science: A Proposal to Minimize Their Use"' [J]. *Veterinary Pathology*, 2024, 61(5): 849.
- [51] Gabovich, A., Kuznetsov, V. 'Eponyms in Physics: Useful Tools and Cultural Heritage' [J]. *European Journal of Physics*, 2024, 45(3): 035802.
- [52] van Raan, J. 'Sleeping Beauties in Science' [J]. *Scientometrics*, 2004, 59(3): 461-466.
- [53] Ke, Q., Ferrara, E., Radicchi, F., et al. 'Defining and Identifying Sleeping Beauties in Science' [J]. *Proceedings of National Academy of Sciences*, 2015, 112(24): 7426-7431.
- [54] Wagner, A. *Sleeping Beauties: The Mystery of Dormant Innovations in Nature and Culture* [M]. London: Oneworld Publications, 2023.
- [55] 杨立昆. 科学之路: 人、机器与未来 [M]. 李皓、马跃译, 北京: 中信出版集团, 2021, 53.
- [56] Söderqvist, T. (Ed.) *The Historiography of Contemporary Science and Technology* [C], New York: Routledge, 2012.
- [57] Doel, R., Söderqvist, T. (Eds.) *The Historiography of Contemporary Science, Technology, and Medicine: Writing Recent Science* [C], New York: Routledge, 2006.

[责任编辑 王巍 谭笑]