

同行评议小史

(中国科学院半导体研究所 姬扬 编译自 Melinda Baldwin. *Physics Today*, 2017, (2):44)

与大多数人的印象相反，同行评议的历史很短、也很复杂。

1936年夏，爱因斯坦投稿给《物理学评论》，论证引力波不存在——但这与科学共识有矛盾。六周后，主编回信并附上一份评议书，请他答复。

很正常吧？我们都知道，科研论文在发表前会被审阅、评议乃至修改。但是，对于文章被送审，爱因斯坦很吃惊、觉得被冒犯了。他回信说，“投稿是为了发表，并未允许你给别人看。我不答复匿名批评。”他再也不给《物理学评论》投稿了。

也许你会想，爱因斯坦是大牌科学家，他觉得自己用不着同行评议。然而，深入考察同行评议的历史，就会发现情况很复杂。1936年，世界顶级的科学期刊通常并不审稿，《物理学评论》也是如此——爱因斯坦此前的稿件(著名的EPR悖论)就没有审稿，尽管其结论显然与量子力学抵触。

这样就不难理解爱因斯坦收到匿名评议时的困惑了。80年后，几乎所有的学术领域，发表文章都要经过同行评议。这是怎么回事呢？

学术期刊审稿的起源

很多文章把同行评议归于同一个故事。1665年，英国皇家学会授权 Henry Oldenburg 创办《伦敦皇家学会哲学会刊》，世界第一份科学期刊。Oldenburg 认为，文章出版前应当征求专家的意见。同行评议就这样诞生了，成为科学出版中始终如一的部分。

果真如此吗？

这个故事出自1971年的社会学研究论文《科学评价的模式》。该文开创性地研究了同行评议，将其发明权归美于 Oldenburg 并暗示，自17世纪以来，同行评议的形式和功能几乎没有变化。

更新的历史学研究质疑了这个说法。Oldenburg 很少就发表问题征询外部意见，他严格掌控着期刊，有时候甚至把自己称为它的“作者”。并没有正式的投稿程序。他出版自己感兴趣的或可能对读者有用的文章，包括实验文章、对他人实验的报道、关于新书的讨论，甚至自己的信件。

Oldenburg 是科学出版史的重要人物，但是他并没有发明同行评议。该荣誉属于剑桥大学博物学家 William Whewell。1831年，他建议皇家学会给予《哲学会刊》稿件以书面报告、并发表于新期刊《伦敦皇家学会学报》，以便促进科学讨论并提供稿源。

皇家学会采纳了建议，但很快就抛弃了出版报告以资公开讨论的想法。学报确实发表过一些报告，但到了1830年代中期，就不这么做了。学会认为，审稿意见主要用于避免发表太尴尬的内容。19世纪中期，《哲学会刊》的审稿工作主要由两位学会秘书管理，分别负责物理学和生物学。评议报告是保密的，只在学会内部使用。很多年里，无论文章接受与否，评议报告都不会

透露给作者。

作者看不到报告，也就没有“修改和再投稿”。《哲学会刊》对投稿非录则拒。有时候也会鼓励作者做些修改。物理学家 George Gabriel Stokes 当了30多年的物理学秘书，经常写信建议作者修改——转述其他审稿人的意见，或者把自己的审稿意见复制给作者。

20世纪早期的审稿制度

19世纪末期，科学界对审稿人的看法发生了重大变化。对科学文献总体质量的担心日益增长，审稿人不仅要维护科学社团或期刊的名誉，更要保护科学文献的整体声誉和可靠性，阻止大量的“污泥浊水混入科学的清流”。

20世纪早期，审稿扩展到英语世界的大多数科学社团。审稿人名义上来源很广，实际则通常出自杰出科学家的小圈子。那时的审稿不像现在这么正式，作者往往也看不到评议报告。

《物理学评论》的编辑会把审稿人的评论转述给作者，审稿人通常递交简短的、非正式的甚至挖苦性的报告。曾经有人建议，某篇文章“如果用隐形墨水来写”就会好很多。1935年，《物理学评论》才提供了标准的审稿注意事项。1960年代，正式要求所有的文章都要审稿。

“冷战”以前，商业化期刊更不可能系统地审稿了。进入20世纪很久以后，《哲学杂志》或《自然》等机构的编辑仍然独立自主。那些期



图1 英国博物学家 William Whewell (1794—1866)在1831年提议,皇家学会应当整理并发表关于《哲学会刊》稿件的评议报告

刊优先考虑出版速度,编辑们觉得,判断某篇文章的好坏,无需咨询顾问小圈子以外的人。

非英语世界的杰出期刊,大多依赖编辑的判断。他们通常把该国最受尊敬的科学家列为编委。普朗克就是《物理学年鉴》的长期编委,没有谁会质疑他的判断力。

基金组织的外部评议与期刊审稿类似。第一次世界大战后,私人基金组织如洛克菲勒基金会通常由中层管理者决定资助与否。与政府或科学社团有关的基金组织,更喜欢外部评议,但绝非常事。

1948年,美国国立卫生研究院(NIH)成立。评估项目申请时,几乎没有外部评议人。NIH自己的专家组进行审议,而资助决定的终极权威是NIH相关研究所的负责人,例如国立癌症研究所。

1950年,美国国家自然科学基金会(NSF)成立。与NIH相比,NSF更看重外部专家对提案的意见。一些提案进行通信评议,其他则由特别专家小组审议。

与NIH一样,NSF领导具有决定权,评议人的看法只是帮助他们做决定——重要,但并非决定性因素。申请者看不到NSF和NIH的评议报告,只能通过摘要得知接收或拒绝的主要原因。

在“冷战”以前,杂志社或基金组织避开审议、由编辑和领导掌握裁决大权,真是司空见惯。爱因斯坦和《物理学评论》的冲突表明,学者习惯了编辑和基金会领导做决定,并不认为外部评议更优越。作者怎么会更信任匿名的评议人,而非真实又可敬的编辑或领导呢?

公众信任和同行评议

1960年代,“同行评议”首次出现于科学出版业。有趣的是,这个说法并非出自期刊,而是描述基金组织和医学界的审议委员会。

“对于不同的人,同行评议的含义不同。”1973年《美国医学协会杂志》社论说,美国医学工作者、英国上院议员和科学家对于“同行评议”的理解是不一样的——根本就提期刊的审稿。在1970年代,科学家和编辑们逐渐接受了同行评议的概念。

1975年出现了对NSF项目的争议,越来越多的人注意到同行评议在科学研究中日益增长的重要性。美国科学家特别是物理学家带着成功研制原子弹的光荣进入“冷战”时期。1953年,美国政府的科学投入比战前增加25倍。1957年,苏联成功发射第一颗人造卫星,科学的公众形象就更高大了。

然而,美国资助科学的热情是有限的。1966年的一份调查报告说,国防部对基础研究的投入没有达到预期目标,例如新式武器的研制。立法者和评论者因而更加质疑

科学的投入。这份报告预示着,科学家在“冷战”早期取得的社会地位和经济地位受到了威胁。

1975年,“冷战”已然相对平和,与苏联争雄不再那么紧要,而美国面临经济危机。为了报复美国支持以色列,几个主要产油国拒绝卖油气给美国。经济增长失速,通胀和失业率暴增。参众两院力图减少开支以应对税收不足,一些立法者盯上了NSF。

NSF最主要的反对者是擅长宣传的参议员 William Proxmire。1975年3月,他开始颁发金羊毛奖,授给最浪费纳税人钱财的政府项目。头两个奖给了NSF项目:威斯康星大学关于人际关系的社会学研究,心理学家 Ronald Hutchinson 关于人、老鼠和猴子为什么在紧张时会咬牙的研究。Proxmire 号召 NSF “别再过家家了”、宣布 Hutchinson 的“瞎扯淡”研究“把纳税人当猴子耍”。

众议员 John Conlan 开始批评 NSF 的教育项目,特别是“人:案例研究”(MACOS)和“个性化的科学指导系统”(ISIS)。MACOS 自1970年代早期就有争议,被指责提倡道德相对主义。ISIS 面向四年级学生,被指责性教育太直率了。

Conlan 要求调查 MACOS 和 ISIS 为什么得到资助,与 NSF 领导层包括主任 H. Guyford Stever 发生了冲突。Conlan 要求 NSF 提交全部评议报告和评议人姓名。Stever 回答说,评议人得到过“隐含的保密承诺”,提交报告文本或姓名违反 NSF 的政策。Conlan 不服,“再次提醒你,我是众议员,也是监管 NSF 委员会的成员……我再次强烈要求你提供评议报告和姓名——原文全本,而非转述。”

NSF项目的公开争论和私下交流，导致了1975年7月的国会听证会，在6天里详细讨论了NSF的同行评议程序。

Conlan说，NSF领导层的决定权太大了，而评议报告没有得到足够重视。NSF必须负起责任，唯一的方法就是公开评议人的报告和姓名。

NSF团队有备而来。Steuer宣布，从1976年1月1日起，申请人将得到评议报告的全文而非摘要。但他坚持，评议必须匿名以保证直言无忌。NSF将建立新的审查办公室，保证领导们认真考虑评议报告，让评议人掌握更大的决定权。

听证会过后，NSF的教育项目显著降格了，MACOS和ISIS的资助几乎完全取消。NSF的同行评议改革暂时消除了激烈的批评，争议从公众视野里消失了。Hutchinson起诉Proxmire、告他诽谤。Proxmire最终公开道歉，此后的金羊毛奖发布会也不再点名道姓了。

批评大多针对社会科学，但所有的科学家都关注这场争议。《今日物理》密切报道了听证会和NSF的政策变化。主编Harold Davis说，听证会表明“如何决定在给定领域内分配资助，同行评议是目前最好的方法。”他还宣布向议员赠送刊物，帮助他们了解科学界的工作方式。“社会事务越来越紧密地与科学和技术交织起来，我们需要更多的同行评议。”

20世纪60和70年代，人们对同行评议的看法发生了重大变化。在20世纪中期，外部评议仅仅是决定录用或资助与否的几种方法之一。到了“冷战”后期，同行评议已经是必要的前提。

NSF的争议表明，美国对同行

评议更加重视的原因是，“冷战”期间科学家和公众的关系发生了变化。基础和应用研究方面的投入，在50和60年代增长迅猛。一旦公众开始怀疑项目的价值，科学家就可能失去公众信任、得不到研究资助。立法者希望获资助的科学家负起责任，科学家则希望由专家做决定。同行评议保证只有最好的和最必要的科学得到资助，也许可以消除分歧。

同行评议处于危机？

今天，为发表文章或得到资助，必须经过同行评议。然而，很少有谁认为这是完美的。有人痛恨错误乃至虚假的结果出现在同行评议的期刊上，有人抱怨同行评议系统过于保守、扼杀科学创新。

2014年，科学公共图书馆(PLOS)的共同创始人Michael Eisen对《华尔街日报》说，人们不应当认为“学术期刊的同行评议表明了科学工作的正确性。这只是说明有几个人(1—4个)读过了，没有发现严重的缺陷。”

同行评议的另一个缺点是，评议人的劳动通常无助于其得到教职或晋级。科学家疲于奔命，无心认真撰写长篇报告，聊尽其责而已。

网络化的出版和阅读，带来了审查文章的另类方法，例如，允许科学家在文章后评论。物理学家早就依赖非同行评议的arXiv.org跟踪最新进展，但他们谨慎对待张贴在arXiv、却尚未发表的文章。

有些杂志尝试略微不同的同行评议系统。PLOS One要求审稿人关注稿件的科学质量，不要操心其重要性和影响力——哪些文章最重



图2 民主党参议员 William Proxmire(1915—2005)在1970年代召开尖刻的新闻发布会，大力批判NSF

要？科学工作者会在出版后判断。eLife让审稿人和编辑彼此通讯，最终对稿件做出综合判定，而不是给出多份甚至彼此冲突的报告。

展望未来，应当回顾过去。我们通常以为，自牛顿时代以来，同行评议就是科学的一部分，几乎没有变化。其实，同行评议的历史很短、也很复杂。同行评议还饱受批评。1845年，审稿人就被认为“充满了嫉妒、怨恨和恶意，一点儿也不宽容。”关于评议无用、充满偏见的说法，历来如此。

起初，同行评议并不是为了发现欺诈、保证科学断言的精确性。Whewell认为评议会刺激科学讨论，科学社团利用评议避免尴尬。文责自负，与审稿人无关。20世纪才认为，审稿人应该对科学文献的质量负责。到了“冷战”时期，只有经过了同行评议，才有科学合法性。

同行评议从来不是一成不变的。根据科学实践的需要——科学社团的信誉，科学文献的质量，或者纳税人的信任，同行评议的形式和目标不断变化。如果要改造同行评议制度，科学家们必须考虑：为什么目标服务？两三个评议人的报告，能否实现该目标？