



科学的不当评价

Peter A. Lawrence

Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

RC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.

E-mail: pal@mrc-lmb.cam.ac.uk

[摘要] 现代的科学社会是一个“审计社会”，广泛采用影响因子和被引频次对期刊和科学家进行评价，但是这种评价方式却有助于科学不道德行为的滋生。该文总结了科学行为扭曲的6种情况，探讨了它们的形成原因，最后尝试提出了解决问题的办法。

[关键词] 影响因子 被引频次 不道德行为 期刊 评价

1 引言

如果用评价科学家的方式来评价歌曲作者，那将会变得非常有趣。假如小调、咏叹调、赋格、托卡塔评估机构的官僚们根据创作的曲目数以及曲目发行后两周内电台的播放情况对歌曲进行排名的话，作曲家很快就会发现要取得事业的发展，不是去做真正的音乐创作而是要创作圣诞歌曲或者迎合顶级电台的流行音乐主持人口味的音乐。然而，真实的科学世界却不像音乐排行那样有趣——片面的评价标准，比如影响因子和被引频次，能够控制人的意志，扭曲人的行为并决定人的职业生涯。

试图对研究数量与研究质量进行评价的行为已经威胁到了现代科学特别是生物医学的发展。科学家也依据这种评价方式进行排名，而排名会影响到经费、职位和晋升机会的获取。这种评价方法起初并未显示太多不足，但是随着它的广泛采用，其缺陷却不断显现。目前的科学界是一个“审计社会”^[1]，科学家的目标和行为都受到了诱导，已将满足评价标准放在了理解自然规律和疾病原理的前面。期刊的优劣用影响因子来衡量，科学家和机构的级别按他们发表论文的期刊级别来评价，这所产生的结果就是，在过去20年中，科学家已经将

主要精力从科学研究转向了撰写论文并设法将其发表于“最核心”的期刊^[2]。现在又有了一个新的评价趋势：根据作者论文的被引频次对作者进行排名。由此，我断定引文“捕获”和引文交易一定会引起人们的广泛追逐。

2 影响因子和被引频次

期刊的影响因子是指过去两年中期刊所发表论文的平均被引频次^[3-4]。尽管论文质量与期刊的影响因子之间有一定的相关性，但是这种评价方式仍有很多缺陷^[5]。需要特别说明的是，期刊的影响因子是一个针对期刊的指标，并不反映单篇论文的被引用情况^[6]。你的一篇文章有可能被证明是错的，或者会白白耗费数百位科学家的努力，但是它仍会是你简历中的闪光之处，也会为你带来一份工作。真正原创性的工作通常需要两年以上才能获得学界的认可和重视——20世纪生物学领域里最重要的一篇文章在头一个10年里罕有引用^[7]；使Ed Lewis获得1995年诺贝尔奖的一篇文章，在发表后的两年里很少被引用，但是在6年以后达到了被引用次数的峰值^[8]。

非常值得注意的是，非常多的引用并不是基于阅读原文本身而是其他论文，特别是综述性文章，故而影响



了影响因子的真实性。一项研究表明，论文后的参考文献只有大约20%被真正地阅读过^[9]。对我们的一篇文章的48次引用^[10]进行分析发现：只有8次与论文的内容密切相关；其他的40次中的3次是完全错误的引用，37次是牵强型引用——实际上应该引用另一篇更合适的论文。因此总体上说，论文的被引主要取决于可见度和可获得性而不是其内容和质量。

然而，目前被引频次却被用来进行科学家间的定量比较。 h 指数^[4]统计了作者署名的所有论文数和这些论文的总被引频次，如果这位作者有50篇论文的被引频次都达到或超过50次，那么他的 h 指数为50。这种评价方式至少与论文本身相关，其中有些论文已经发表了多年，也可能发表于影响因子较低的期刊。

采用 h 指数和影响因子来评价科学家的方法使不道德行为进一步加剧，它能够使挂名作者获益。挂名现象非常普遍，甚至已经成为常规，很多挂名作者甚至对研究内容一无所知。论文挂名有几种情况：或是提供了反应试剂，或是研究组负责人间进行了交易，或是权利滥用——包括对年轻科学家的非法利用^[11]。例如，小组负责人在由下级成员起草、组织并完成的方案上署上自己的名字。当一旦出现造假或者错误的情况，挂名者也会受到声讨，但是他们却会以没有亲临实验现场而加以搪塞。耐人寻味的是，他们却很少会因为没有亲临实验现场而拒绝科学奖项。

3 科学行为的转变

不幸的是，前面提到的评价方法对科学家的认识和行为产生了不良影响。我并不是说科学已经完全走向腐朽，只是在下文中描述了科学行为的趋势和极端现象。很多研究者和教授是具有原则的，但是他们面临着更加困难的环境。

首先，科学研究的本质受到影响。在高影响因子期刊发表论文是至关重要的，因为仅仅一篇这样的论文就能够改变一位博士后的前途，2~3篇这样的论文就能够确保稳定的事业。这样的现实已经严重扭曲了科学思想，科学家已不再将精力倾注于科学问题和方法的探索，而是倾注于投稿、审议和出版的流程。现在的研究者发表论文大都经历这样的过程，先向高影响因子期刊

投稿，然后再向低影响因子甚至更低影响因子的期刊投稿。通常情况下，科学家需花费一年的时间来等待论文通过编辑、审稿；如若被拒绝，则会修改论文以避免更多的批评，并想办法提出各式争辩的理由，然后会恳求编辑寻求一位更合意的审稿人。如果这样的策略在此种期刊不能奏效，那作者会投入到下一种期刊的战斗中去。这对于作者的时间和精力是极大的消耗，但是却能够带来事业上的回报。因此，我希望经费主管部门能够调查其所资助的研究组负责人在论文经营上所花费的时间和努力。难道论文第一次就投送到有可能接受它的期刊不是更有意义吗？新近大家形成了一种看法：论文发表过程像擂台赛一样充满了艰辛和挑战。

第二，科研人员为迎合评价标准而改变研究策略，避免承担太多的风险：长期进行新事物的研究有利于产出创新成果，却不利于获得经费的持续性支持。研究要跟风并且要保持与外界联系，这样至少能博得他人注意；反之，如果涉入新领域的研究，将会面对孤独无闻的风险。一篇新领域的论文并不会使编辑眼前一亮，因为他们清楚很难找到合适的审稿人。在顶级期刊发表的任何医学论文都能够获得大量的被引，因此，假设你的工作与医学相关，那么由于前述原因你也会选择最热门的研究方向，因为发表关于人类的肤浅但热门的研究结果将会比发表关于斑马鱼的深刻研究成果要容易得多。

第三，研究结果的展示方式：夸大研究成果，尽可能地分解研究发现（按现有评价方法4篇论文肯定比2篇论文好），压缩研究结果（多数顶级期刊版面有限，例如*nature*的“通信”栏目版面密度如同“黑洞”一般），简化结论但却将研究材料复杂化（这样审稿人将难以挑错），包含并挖掘大量的数据源（即便它们缺乏新意）。其中最负面的现象是，忽略或掩盖那些不符合论文主题的结果。一系列这种做法，往往将论文变得凌乱而无吸引力。

第四，研究开展和论文署名的方式。前述评价方式促使研究组向更大规模方向发展，这是最简单的计算题，因为研究组负责人会在所有的论文上署名，因此，研究组成员越多，就有越多的论文发表。如我猜想的，在大的研究组中年轻研究人员如果失败的话，并不被记录下来，并且没有人考虑葬送了多少人的梦想，浪费了



多少人的生命，这些人的失败也不会使得研究组负责人显得低产。我相信追求论文数量也致使学生被当作技术工人来对待；于是学生只能言听计从，却并没有学会如何成为一个真正的研究人员。同时，研究组负责人经常将其学生的研究工作发表：如果他并不十分明确研究结果或者不能确知如何获得的结果，那么他将可能对数据进行大胆的解释。如果将来研究成果能发表在高影响因子的期刊上是他们的福祉。这样所带来的问题是，论文可能缺乏真实性，但这是不易测定的。可以肯定地说，学生没有获得恰当的教育，并且会发现博士生活枯燥无比；可是这可能对导师有利，因为学生继续从事研究并成为竞争对手的可能性将会减小。

第五，科研人员离开实验室转而去建立人际网络，这将有助于与同行甚至论文审稿人之间建立相互协作的关系。通过参加大量的会议、进行访问交流和维持人际网络将有助于加深研究者在编辑（经常参加学术会议）心中的印象，和编辑建立关系有助于研究者论文的评审。（在顶级期刊中有90%~95%的稿件被编辑拒绝，但是30%~50%的送审文章被最终接受）。因此，毫无疑问一些最成功的科学家会花费大量的时间进行人际网络的经营。

第六，不当的评价方式还会产生其他不良影响^[12]。现代科学的生存竞争、竞争的公开性以及部分“张杨者”具有的竞争优势使得那些儒雅并谦和的人处于劣势；实际上并没有证据说明缺乏进攻性的人缺乏创造力。因为多数女性不具备进攻性^[13-14]，所以，尽管参与生物医学研究的女学生越来越多，但是顶级有影响的女性研究者却没有明显的增加^[15]。性情温和者能够确切感知排挤他们的人的存在^[16]，他们会以行动选择离开这个行业。我认为我们不仅丧失了大量的有创新能力的研究者，而且科学在充满理解和更和谐的环境里会取得更好的发展。

4 问题的根源和解决办法

风气是问题的根源，目前的管理模式和政策导向都偏重于采用数字方式对成绩进行评价，然后实施奖励。截至目前，管理者及一些其他人都忽略了科研管理机

构存在的根本目的，其内部的运行已经变得越来越官僚化，并且是以研究成就作为代价的，例如，增加了经费申请和研究评估的复杂程度。这种官僚化对科学家有过重的要求，这摧毁了他们的意志，挫伤了他们的自尊。各个层级的科学家们都在批评我们对这种情形的不作为。我们应该对该问题保持警醒，并尽早做出改善。例如，当聘任委员会在决定人员的雇佣时应该时刻意识到他们是在雇佣一个具有各种能力的人而不是一个数字，而能力中最重要的是研究中的创新能力。我不认为创新能力应该用影响因子或者被引频次来衡量，有人也赞同我的意见（见<http://voltaire.members.beeb.net/goodscience.htm>）。目前，集中评价有助于国家范围甚至世界范围的比较，但是这种评价方式被高估了。在聘用、晋升、确定终身职位时应该有更多的候选人，比如可以通过阅读他们的著作、聆听他们的演讲等方式加以确定。

为了提高行为标准，应该制定一套道德规范和一系列实施措施，特别是在出版方面。目前，人们并没有对出版过程的恰当行为做出界定，例如，署名的问题。我觉得署名问题的关键是知识产权的归属。（例如，我的学生在高影响因子的期刊上发表论文，而我并没有署名，结果我所在学校不能将其用于科研评估）。关于如何使署名更具公平性的公开讨论是一个良好的开端。稿件的审稿过程仍有一些缺陷，一些匿名审稿人会因利益而扼杀论文，他们会利用自己的信息特权向他人透露论文的内容，这些现象应该得到重视。作者如果感觉受到瞒骗也没有别的办法，只能寄希望于下一个期刊。可是这段延误的时间对论文成果却有致命的影响，甚至有可能影响作者职业前景和获得的支持。即使作者感觉到审稿人和期刊对待他们的论文有所不公，那么他们又能做什么呢？目前，期刊可以先接受一篇论文，或者要求作者花几周的时间进行修改，再确定一个出版期限，然后发送校样，但最后却撤掉了论文（我与Science之间发生过这样的事情）。解决这个问题的一种可能途径是在经费管理部门建立调查制度，使那些受到不公正待遇的作者有机会提出申诉。应该是让身处第一线实干的科研人员获得支持的时候了。



参考文献

- [1] Power M. The audit society: Rituals of verification. Oxford, UK: Oxford University Press, 1997.
- [2] Lawrence PA. The politics of publication. *Nature*, 2003, 422: 259-261.
- [3] Garfield E. How can impact factors be improved? *Brit. Med. J.*, 1996, 313: 411-413.
- [4] Hirsch JE. An index to quantify an individual's scientific research output. *PANS*, 2005, 102: 16569-16572.
- [5] Seglen P.O. Why the impact factor of journals should not be used for evaluating research. *Brit. Med. J.* 1997, 314: 497-511.
- [6] Editorial. Not-so-deep impact. *Nature*, 2005, 435: 1003-1004.
- [7] Olby R. Quiet debut for the double helix. *Nature*, 2003, 421: 402-405.
- [8] Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature*, 1978, 276: 565-570.
- [9] Simkin, M.V., Roychowdhury, V.P. Read before you cite! *Complex Syst.*, 2003, 14: 269-274.
- [10] Casal J, Struhl G, Lawrence PA. Developmental compartments and planar polarity in *Drosophila*. *Curr. Biol.*, 2002, 12: 1189-1198.
- [11] Lawrence PA. Rank injustice. *Nature*, 2002, 415: 835-836.
- [12] Lawrence PA. Men, women and ghosts in science. *PLoS Biol.*, 2006, 4: 13-15.
- [13] Geary DC. Male, female: The evolution of human sex differences. Washington, DC: American Psychological Association, 1998.
- [14] Baron-Cohen, S. The Essential Difference. London: Allen Lane, 2003.
- [15] European Commission. Science policies in the European Union: Promoting excellence through mainstreaming gender equality. Brussels: European Commission, 2000.
- [16] Babcock S, Laschever S. Women don't ask: Negotiation and the gender divide. Hoboken, New Jersey: Princeton University Press, 2004: 240.

The Mismeasurement of Science

Peter A. Lawrence

Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

RC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.

E-mail: pal@mrc-lmb.cam.ac.uk

[Abstract] Modern science is an audit society. Impact factors and citations are broadly used to evaluate the journals or scientists, but the measures have caused some unethical behaviors. The paper summarized six types of behavioural change and analyzed their reasons. Several approaches were pointed out.

[Keywords] impact factor, citation, behavior, journal, evaluation

翻译: 李志忠