

期刊同人评审体制的目标、过程和价值

姚远 穆军 吕时群 编译
孙宏 审校

提 要: 本文运用统计数据、实际工作经验和前人的研究成果,对同人评审体制的目标、评审人的赞同及偏见、评审过程和目的、评审过程的价值,以及编辑、作者和审稿人在这一体制中的关系等进行研究。本文将期刊的同人评审体制定义为:“由专家(同人)对提交科技期刊发表的材料作出评价。”文中概括了同人评审功能的四种形象比拟,即筛子、转换器、锤子(同人评审)和铁砧(编辑标准)、处在黑暗中的射击手。本文认为有关期刊同人评审的大部分研究在方法论方面是较弱的,并且大部分集中过程的研究,而不是对结果的研究。本文预测将开发一种用于期刊同人评审过程的研究程序,并且会在10至20年间对于大部分关键性问题取得进展,而这种程序对发展整个科学社会的一种相互信任和支持的关系,将是必要的。本文还认为期刊同人评审体制是一种重要的研究客体,呼吁加强这方面的研究,并将其列入某种研究的议事日程。

0 引 言

对于许多调查研究者来说,期刊的同人评审体制是一件很生疏和很神秘的事情。关于对投往期刊发表的科学论著所作的同人评审,目前流行的看法大致可以用四种比拟来概括:一是筛子(同人评审将所投稿件中价值者和无价值者区别开来);二是转换器(一位的执著的作者无论写了什么文章,最终能够将它发表,而同人评审则决定他能在什么地方发表);三是铁匠(论文在同人评审的“锤子”和编辑标准的“铁砧”二者间被猛烈锻打至新的和更好的形态);四是处在黑暗中的射击手(同人评审基本上是无法预言和无法复制的,因此其效果是毫无规律的)。把这四种比拟分别来看或者联系在一起,它们在多大程度上可以抓住同人评审的真正本质呢?

虽然基于同人评审的个别决定可能遭受有力的批评,但是整个评审体制却为人们所接受,其运转也相当令人满意。学术上的报偿大多依整于在进行“同人评审”的期刊上发表文章的数目,国会委员会被告知,编辑在同人评审之后采取认可态度,这对于作者和文章两方面都标志着一种荣誉,而新闻工作者也从中学到了同样的教训。政府管理机构已受到被管理行业方面的压力(至今仍无成效),要求这些政府部门接受在同人评审的期刊上发表的作品,以代替由他们自己内部评审的作品。

人们也许指望研究事业的这样一个重要部分的功能会成为一个主要的研究对象。事实却并非如此。极少有扎实可靠的科学论文通过同人评审得以发表,除非特别加以努力,一些重要的空白不可以很快被填补,或者以一种在技术上可能接受的方式被填补。^{[1][2]} 这对我们来说,似乎有些自相矛盾:在科学报告被提交发表时,严肃的、注重质量和富有革新精神的仲裁者却不把审定其他人文章时使用的标准应用于自己

的著作上。

在过去的10年间出版的12篇关于期刊同人评审体制的研究论文都具有下述每一项特征。这些论文都与生物医学科学有关(虽然未必写的是关于这一科学的问题)。这些论文预定对一个特定的假设进行测试或对一个特定的问题进行测验,并且基于一个明确限定的样本。其中没有任何一篇论文在期刊中抽用随机性的样本,并且一个领域的期刊样本可能有很大差异,使得研究结果的普遍性界限不总是很清晰。

1 期刊同人评审体制的目标

我们将“期刊同人评审体制”定义为同专家(同人)对提交科技期刊发表的材料作出评价。这样的一种评审体制,被用于许多目的,并以多种方式影响作者、期刊读者、评审者、编辑和他们的期刊,以及整个科学领域。一个合成性的目标也许即是,当该系统工作正常时,它能:(1)筛分出那些蹩脚的设想、蹩脚的设计、蹩脚的实施、价值不太大的、抄袭的或不可思议的研究报告;(2)保证对其它有关研究工作有适当的考虑和承认;(3)导致对所评定的手稿进行有益的修订并因之使其质量得到改进;(4)帮助指导研究成果投到最合适的期刊上去;(5)通过增进训练、教育和对从事研究的科学家的鼓励,提高整个领域的技术质量;(6)在各篇论文上注出质量的标记,以便有助于可能使用这些成果的非专家人士;以及,(7)对于妥善使用同人评审系统的期刊要增进职业方面对它们的接受和认可。

此外,我们审阅^[1]和讨论^[2]了少量关于同人评审在筛分出没有发表价值的材料(上述目的1)中同人评审的功效方面的文献,以及数量更少的关于文件登记次序方面的文献(目的2)。目的3至7一直是许多评论和推测的一个主题,但似乎并未成为任何科学研究的题目。对于目标缺乏(明确的一致意见,这些目标有时可能相互冲突,以及对于那些最有可能已被广泛地采用的目标缺乏)了解,这都表明存在着一种关键性的持久分歧。总之,我们仍然不知道什么是一个“好的”评审过程的结果,或者除了通过编辑和评审人明白确定的标准之外,应该如何衡量它的业绩。于是大多数这方面的研究集中在过程上(比如同行审稿人赞同),而不是结果上。

2 关于同行审稿人赞同及其偏见的研究

通常假定,通过两位或更多的审稿人一致同意这一个标准即可对手稿的质量作出评价,虽然故意选择观点类似的审稿人也许能通过减少在手稿中检出缺陷的数量而产生相反的结果^[4]。我们将更进一步假定:故意选出对论文持不同看法的评审人。当他们完全赞同时(在他们的理由中,以及在他们的总结性评价中),除了其中一篇之外,其余一切都是多余的,除非是用这些评价使编辑信服那些批评是正确的。人们不能通过对评审人赞同的研究来对同人评审在达到上述7个项目中的任何一项上面是否成功作出鉴定,尽管这样的研究具有其他用处。

谁是同人,或者应该把谁看作同人?关于这个问题我们只能找到少量研究^[5],而且没有人对评审人由于专长范围或能力水平不同而产生的不同意见作过研究,尽管英哥芬格尔(Ingefinger FJ)^[6]对一些专长差异曾经起到重要作用的具体事例作过评论。

在评审人中间过多的赞同可能并非好事,或者至多是中等的,但人们也许可以希望编辑们关于发表还是不发表的决定是具有相当稳定性的。对于经验主义的研究来说,这显然是个棘手的问题。彼得(Peters)和塞西(Ceci)报道了这个问题,但有些作者和几位评论家认为他们的论著有着较大的局限性^[7]。我们知道,在编辑决议方面没有其它系统性的直接研究。

为什么可能有两名专家对于手稿的价值产生不同意见呢?一个因素也许是偏见——即在形成评价中和

对编辑的忠告中不是使用严密的科学和技术的是非曲直这样的标准。偏见可能通过这样一些事情被引起,比如:评审者对作者或单位某些特点的反应;对于符合审稿人自己的观点或普通性观点的结论的反应;审稿人之间在专业范围或科学信念方面的差异;能力或努力的不同水平;等等。

为防止误解应作两点说明:在这种意义上和这种上下文中的偏见是否总是坏事并无定论,对偏见和客观的评价二者如何区分,见多识广的评审者并不总是持相同意见。研究质量不是接受或拒绝一份手稿的唯一合法标准。一名编辑不仅要注意研究工作本身,而留意象对这一研究进行报告的质量这样的问题;注意使期刊覆盖的各个次级领域之间保持平衡;注意一个题目在其他场合或在新近发表过的论文中是否已被讨论过;并注意编辑自己对“是在那里起作用”的感觉;或者读者会很快需要知道什么。给“研究质量”本身下定义是极为困难的,而任何与作者品质相关的偏见的严肃讨论,必须使用某种对质量的界定。

对稿件采用率的三项分析^[10~12]说明了作者名望的水平,他们的年龄和合作者的数量等因素中的每一项都与能否成功地发表文章密切相关。这三项研究的结果没有做到,并且不可能做到说明偏见在工作中已是一种普遍的作法,因为这些论文一点也没有涉及对于质量的评定。这可能是正常的,例如,那些功成名就获得大奖励的科学家会更有可能比其他科学家写出新的和发表价值的论文来,并且由两个或更多的头脑(就象在多作者中已表现出的那样)在一般情况下,要比一个头脑想出的论文更好些。

目前,几种期刊试图在审稿程序上实行对审稿人隐匿作者身份的制度,以减少偏见的机会,既不袒护作者也不与其作对。类似的通常作法是实施对审稿人匿名——不时地招致批评,常常是对其“公正性”的评语。对这两种匿名审稿形式的随便哪一种类的效果,是否存在全面的和客观的评价我们都不了解,甚至不了解隐匿作者身份的作法是否有效。

另一个导致审稿人中不同意见的因素是编辑可能未搞清楚,或者审稿人未能理解,到底应对哪种属性进行批评和使用何种标准。看来,同人评审所用标准的性质在科学和技术领域中存在本质的差别。^[14-17]有一种方法可用于审查那些标准,即对实践进行调查,或许是通过评定一份质询事项表请审稿人在他们作出推荐和批评之后即刻填写出来,而不去调查那种更为概略的和具有自觉意识的观点,这种观点可能在编辑对自己评判手稿中的长期感受经过深思默想之后才会表面化。

3 过程和目的的研究

这种体制的规则清楚了吗?和同事随便谈几次话便可了解到编者、作者和审稿者相信,他们对期刊的评审过程应当怎样工作有着颇为一致的理解,尽管严重的失误可能频频发生。我们不能确信在一个彼此了解中的这种信念是正确的。

例如,有一个问题是,需要什么样的同人评审才能使一份期刊具备同人评审刊物的资格。象《柳叶刀》(The Lancet),那是普遍受到尊敬的已具有类似同人评审体制的期刊,它仅有一小部分稿件提交外审。莫格芬格尔^[6]暗示或许一些著名的期刊会省去同人评审,但在大多数其他期刊中,反而却会长久地存在下去。

贝耶尔(Beyer)^[17]发现了在科学领域中投稿和已审稿在比例上的差别。许多编辑,特别是那些著名期刊的编辑,会收到一部分小比例的带有非科学内容的“狂热者”的稿件,也会收到一些比例的实际上毫无技术性优点的稿件。对于此类材料作即时拒绝,是毫无异议的。然而,在论文未能达到编辑的一般性标准方面看法可能有分歧,例如与期刊的宗旨相关与否、最低限度的语法和文体标准的胜任程度、新资料或新认识的质量、长度,以及主题内容等。对此类论文不经审阅便予立即还可能为作者及审稿者提供服务,但关于此类论文范围划分界线始终不清楚。

我们自己作为评审人以及常常作为对其他人评审的读者,我们都在编辑的瞩望中和在给予评审人的指示中已发现了相当大量的不明确表达。这种不明确的一部分是蓄意造成的;几名编辑告诉我们他们的忧虑,对

审稿人过分详细的指示可能会约束潜在的有用评论。似乎没有证据表明使用这个还是另一个指示可造成多大差别;的确,我们怀疑许多审稿人根本不会阅读这种指示。

在决定一篇特别的稿件发表或退稿时,即各个审稿人有一种一致意见,也可能被编辑所忽视。有某些限度的证明^[13],认为编辑在他们对审稿人的及时推荐的运用上有所变化。这种许可为编辑及审稿人提供了一些应付的余地。当然,如此“偏见”不一定是坏的,而假如一名编辑是确切有力地在引导本领域进步的话,它可能是必要的。由于“编辑的”原因,而非科学上的原因审稿人可能建设采用或退稿,而编辑必须对这些原因作出判断。我们不知道其他在这方面进行了调查研究的工作,即使在描述的水平上,也未见有人研究过编辑在多大程度上利用了审稿人的即时评论以及以何种方式利用它他们。

总的说来,一些审稿人相比之下可能更难满足其要求。所有编辑似乎很可能认识(有些还利用)一贯持赞同的看法或一贯持否定看法的审稿人,尽管史密格爾(Smingel)和罗斯(Ross)^[14]未能找到证实这种意见的根据。一份物理学期刊用一个“凡有疑问则接受”的总政策,而大多数情况下将同人评审用于常常会被最终退稿的论文。这或许是为了通过书面理由保护编辑,更确切地说不是为了记载不同意的理由而不是作出最初的退稿决定。

评审过程在选择和改进重要的著作给以发表以及拒绝粗劣的或非重要的著作方面的功效几乎统统没有文献记录。然而,可资利用的资料表明审稿人的评价和继论文之后的引用这二者间的关联甚少。^[6,13]在评价已发表著作的质量或重要性上,有许多关于与该文章被引用次数相关联的问题,不过这样的引文次数也具有其他方面的用途。

我们仅发现两项稿子有多少次被投到这份期刊却被另一份期刊采用;或者在这一过程中它他经过了多少修改的主要研究^[19,20]。威尔逊(Wilson)^[19]报道有85%被《临床诊断杂志》(Journal of Clinical Investigation)退回的手稿随后又在别处被发表。其中有1/6包含有进一步的加工,但报道未说明有多少篇稿子是因为需要进一步加工而被退回的。罗曼(Relman)^[20]报道了值得注意的有关《新英格兰医学杂志》(New England Journal of Medicine)退稿情形的类似数字。

对于研究中存在不道德实践的程度或者同人在维护道德和职业标准上所起的作用,我们不知道在这方面进行过任何研究,无法决定这两者程度。然而,以前一个参议院小组委员会^[21]关于新药的商业性研究已揭露出与此相关的大量辩论,并且有关好的评审者应如何察觉如此陋习的研究,亦有少许可得到折资料。投稿到期刊与起改进作用和工业试验性的问题相比,常见的道德问题可能发生的更少。在其他方面,对国会揭露的问题来说,评审者并没有得到强有力的有效控制,象表明在誓约之下成为证据的问题、传唤的权限、顺序的运用和可疑的审查者,以及发生被公众质询的非难和反驳的机会等等。几个新近的事件表明,彻底的作弊在某种程度上侵袭着科学期刊,并且这样的作弊可能会难以察觉。^[22-24]

4 评审过程的价值

评审体制最主要的价值意味着专业人员的时间、在发表中的延误、在定稿质量上可能存在的退化变质、对可能象征着一种科学进步的文章的拒绝,以及信誉的丧失。罗曼^[20]估算,在历年的《新英格兰医学杂志》的评审过程中,需要6~7人/年的同人评审,2人/年机构内部工作和10万美元(1978年美元)的办公开支。英哥芬格尔^[6]说过:“评审体制以其大无比、不加咀嚼的巨口吞食时间”。我们只有极少量关于这些总数的精确资料,但金等人(Kin et al)^[25]估计审稿人用于审阅每篇稿子的时间均为6小时。

温(Wong)^[26]提出过分地批评审稿人和编辑,会导致作者采取某种偏见。比如,不报道前后矛盾的结果和定性的结果、省略那些已受到批评的方法的详细参考文献和只给出主要的和含糊的过程叙述。菲杰(Feige)^[27]提出一些由许多事件和只报告正面结果构成的研究分析数据。这种现实是含糊地超出科学伦理观的正确

标准的,但它通过发表了虚假正面结果的增长数,歪曲了科学文化。泽论(Zelen)^[29]曾评论了这方面的一类问题。

通常价值的另一种类是联系到同人评审体制通过评审者不适当地给某些题材以优惠。我们找到了有关这种情况^[6,29]的少数轶事的报告,但那是唯一未被确认的一个普遍问题的继续存在而已。

削弱同人评审的价值无论是否可能,那也许是在方法上的一些艰难尝试和正当探索,通过审稿人,同人评审将会被运用得更具优势。唯一对编辑、作者和可能接受的审稿人来说,是直接受益于评审的一般自然增长;对读者来说是非受益者,只有可能在编辑的决定中对用稿还是退稿带来影响。几种期刊发表编辑过的一些论文的评审意见。这种实践是有效的吗?它能起到改进的作用吗?它将会持久下去吗?有没有其他的方法使对评审者成果的和利用变得更好呢?

5 结论

有关期刊同人评审的大部分研究在方法论方面较弱,并且大部分集中在关于过程而不是结果上。已发表的工作大部分涉及有关心理学和相关的训练;其余部分大多为已完成的设计和实验,象在科学上社会上的研究,更确切地说是在技术交流方面的研究。

我们相信某些团体或机构,政府的或民间的,将会着手人做一种非常精确的期刊同人评审功效和相应研究所必须的分析。一种预测可以很好地证明,对开发一种研究程序并用于期刊同人评审过程中的研究是必要的。这样一种程序不一定是非常昂贵的。对“应急”计划没有明确的需要,尽管我们相信尽快着手制定这种计划。有10至20年时间,六名资历较深的调查研究者再加上某些辅助人员,并不需要别的财力物力则能在大部分关键性的问题上,取得重大进展。

这种程序对于发展整个科学社会的一种相互信任和支持的关系,将会是必要的。编辑们将会需要这种全心全意的合作。假如许多重要的调查研究已经找到了在伦理研究上折正确标准,那么来自于审稿人和作者的有资料依据的赞同将会是必不可少的。程序的公认应通过进行学术研究的科学家和期刊读者提供实质性贷款资助上经历一种社会事业性的利害关系。

有可能做些什么呢?我们引述以下可能发生的事例。同人评审的价值可能涉及到各种各样的、易于证明是同一特征的投稿,不管是否看出论文大量的充裕,都会迅速地处理和恰如分的外审。一项横跨各期刊的全面评述可能研究对审稿人隐匿姓名和反过来对作者隐匿审稿人姓名的功效好处和不利条件。稿件的一种鉴定后修正可能被自始至终地跟踪,对于调查“筛子”和“转换器”二者的同人评审外貌来说,这是出于促进这种作用,以及减少花费的目的,借以对“筛子”和“转换器”的同人评审外表进行调查。那种下达给审稿人的不同种类指令的功效,可以加以比较。其性质、限度、价值和改变建议以后的效果可通过审稿人加以比较,这种比较既包括那些退稿和投往别处的稿件。或许大部分有价值,技术质量的客观估量和科学信息容量的价值,在稿件中也许已被发展和被证实,以致审稿人的评语、编辑的决定、对读者价值和质量的各个方面这四者之间的关系,会得到更好的理解,并最终得到改善。其他研究题目很丰富,不再赘述。

最大的直接需要是对于领导的需要。期刊同人评审体制的研究已引起大量的忧虑,我们认为这是因为那个方面存在着些许高水平的兴趣,已经发现了此类工作方面不易认识的原始资料和已经发现了未被理解的所需的公开分析。结果,我们已发现这一工作常常是蹩脚的设想、衰弱的方法论、小规模尝试的倾向、没有一个较重要的个人承担对期刊同人评审体制研究或相关事情的长期赞助,并且策略也是离题的。我们并不归咎于调查研究者;总的来讲他们值得给予正确评价,就象拓荒者在艰难的境况下所可能做到的那种最大努力。现在我们的结尾视其为更感兴趣的事情,并且清楚地认识到期刊同人评审体制是一种重要的研究客体,它可以导引改进期刊在推动科学前进方面的功用。

尽管我们作了许多批评性评论,我们却对期刊同人评审给予了高度的注意。虽然我们认为某些业务上的细节也许已获得真正的改进。但我们却不知道对体制本身有什么切实可行的选择。