

中国的物理论文

——兼谈科学论文的评价

王鼎盛[†]

(中国科学院物理研究所 北京 100190)

改革开放三十年来,我国的物理学研究有了明显进步.这些进步反映在各方面取得的业绩,包括在发现物理现象和探索物理规律上的贡献,解决国家发展的关键问题中发挥的作用,物理研究的基础设施的研制和建设,以及物理人才的培养和向其他领域的输送等等.

作为反映这些业绩的一个重要的标志,是我国所发表的物理论文数量上的增多和质量上的提高.科学论文的作用不但反映在它是学术交流的重要平台,更加重要的作用是在科学知识的积累上,科学论文是不可替代的最重要的载体.考察和分析一个单位或者一个国家所发表的物理论文,考察和分析它们在科学发展中产生的影响和作用,可以大体上了解它们在物理研究上的取向和研究能力,评价它们所做的贡献.近三十年,尤其近十多年来,我国在物理研究领域发表的论文不但数量有大量增加,而且还有一些突出的成就取得国际同行瞩目,例如上世纪八十年代末关于氧化物高温超导的研究,以及2008年以来关于铁基超导体的研究等等.

经过一段时期,历史地去分析一个国家发表的物理论文,可以看出它的物理研究的发展轨迹,也会反映出存在的一些问题.本文作者根据自己对国内物理界的了解,与物理界同仁交流切磋时所得的启发,再考察历年我国发表的物理论文的有关数据,相互证辨,提出几点对我国物理研究现状的观察与思考.与之相关,还讨论了在我国的科学论文评价中存在的一些问题.限于个人所见,难免不当乃至谬误,冒昧成文发表,不过抛砖引玉供学界同仁讨论之意.

本文在“中国”或“我国”项目下列出的数据,仅指我国大陆(境内)的研究单位或作者,不包括港澳台(境外).大陆与港澳台虽有交流,但由于各地的体制与环境不同,对科技发展设定的目标和要求各有差别,发展历程和存在的问题也不完全相同,不必放在一起讨论.

1 中国物理论文的数量

使用 SCImago Research Group 所做的期刊比较平台(SCImago Journal Rank),可以对各个国家和地区的论文状况加以评估.它使用Elsevier B. V. 公司的 Scopus 数据库,这个数据库收录了全世界的主要物理学和天文学期刊 450 余种.近三十年,中国作者所发表的物理论文,已大量进入了这个国际交流的圈子.到 2005 年之后,中国年产论文总数已升到第二(图 1),仅次于美国.到 2006 和 2007 年,已明显地远多于第三和第四(德国和日本),与美国也相差不远.而且,我国的物理论文总数还在继续增长中.

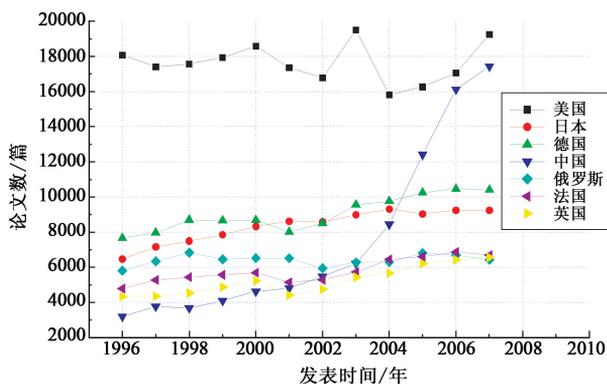


图 1 1996—2007 年,在物理和天文学领域发表论文最多的几个国家的年产论文数比较(数据来源:SCImago Journal Rank, <http://www.scimagojr.com>,包括全世界物理和天文学的 450 余种主要期刊,中国仅指中国大陆)

我国论文数快速增加,主要原因是我国的研究机构和高等学校近十年来大量增加了招收的研究生

2009-10-22 收到

[†] Email: dswang@aphy. iphy. ac. cn

表 1 2000—2006 年中国大陆作者独立发表(无境外合作者)的物理论文数,按总引用次数分档排列.

数据来源:Web of Science;刊物:所有 SCI 物理类(约 260 种);总引用次数累计到 2009 年 9 月截止

影响档	很弱	弱	中	强	很强	极强	论文总数
总引用	0—2	3—9	10—29	30—99	100—299	≥300	
2000 年	2783	1360	701	206	22	2	5074
2001 年	2539	1567	865	212	18	3	5204
2002 年	3045	1864	988	272	26	1	6196
2003 年	2979	2185	1249	306	30	0	6749
2004 年	4057	2683	1388	274	14	1	8417
2005 年	4831	3351	1514	266	16	2	9980
2006 年	5129	2702	852	90	4	1	8778

人数.各单位每年招收的硕士生、博士生数量都大于,甚至远大于其拥有的教授(或导师)的数量.学生最多的教授每年招 5 名以上,甚至有同时辅导着超过 30 名研究生的;几乎所有单位,都对学生规定了发表论文数的额度,通常是硕士 1—2 篇,博士 2—3 篇;对有论文发表,尤其是在著名刊物上发表论文者给予奖励,或优秀学生称号或现金等等.有了多招生、定额度、和给奖励这三个办法,论文的数量增长当然不难立竿见影,节节向上.当前的这种规模扩张,从培养人才的数量上看,应该说不无可取之处.是否就一定可以说成是中国的物理研究的能力,或者对物理学的贡献在不断增长呢?下节讨论我国物理论文的影响,对这个问题得到的答案并不乐观.

2 中国物理论文的影响

正确地衡量论文的科学价值是一项困难的工作,代之而起的是通过考察论文被其他论文引用的情况(数量和评价),去衡量论文在科学界产生的影响.这有一定的道理,因为论文被其他论文引用的情况,标志着论文受到关注的程度,也是论文被同行肯定或否定的记录,基本上如实地反映了论文在科学知识的积累过程中所起的作用.从实际操作层面上看,在有完备的数据库,而且通讯发达的现代,引用情况的检索也较容易做到:收集论文的引用数,只要一点鼠标就可以做到,完全无须任何专业知识;即使要收集论文在引用中所得到的肯定否定或其他评价,也只要有基本的专业知识,花少则几十分钟多不过几个小时就可以完成.虽然衡量或比较不同学科之间的单项研究成果,可能误差很大,甚至毫无意义.但在同一学科里,

作大样本间的衡量或比较,比如研究所乃至国家间的总体比较,该有较高的可信性.这是本文考察中国物理论文的影响时所用的方法.

表 1 中列举了发表时间为 2000—2006 年,发表在 SCI 数据库中约 260 种物理类刊物上,作者仅为中国大陆(即无境外作者合作)的论文的引用统计.对各年发表的论文,均按多年累计的被引用的总次数(统计到本文成文的 2009 年 9 月截止),分成“很弱”到“极强”等六个影响档次,列出各档的论文数目.

表 1 中把论文的影响从“很弱”到“极强”分为六档,是按被引用的总次数以对数尺度划分.为便于不同学科间比较,也有一些研究采用按相对比例划分影响强弱,比如把引用数居于某学科在全世界某年发表的论文前 0.1% 列入该学科当年的“极强”档,前 1% 列入“很强”档等等.本文后面只讨论物理学,估计本文的读者也多为物理学家,他们了解自己论文的引用情况,并不了解他人的引用情况.采用引用次数划分,使这些物理学家很容易确定自己所处的位置.其实,全世界每年发表物理论文十多万篇,引用次数大于 300 或 100 次的大约分别为 100 或 1000 篇左右,也就是前 0.1% 或前 1%.所以这两种强弱分档是大体相当的.

可以看到,中国大陆产生的论文列入“强”影响档以上的论文(累计被引数 30 以上),每年不过 230—340 左右,数目不大,比例较小,占论文总数的约 5%;能列入“很强”影响档的(累计被引数 100 以上)论文每年只有 20—30 篇,只有约 0.5%,可谓凤毛麟角;而列入“极强”影响档的论文,仅只聊胜全无,每年只有 0—3 篇.

还可以注意到,统计到发表后的五年以上时,每

表 2 中国大陆作者与境外作者合作发表的物理论文数,按总引用次数分档排列.

数据来源:Web of Science;刊物:所有 SCI 物理类(约 260 种);论文发表时间:2000—2004 年;总引用次数累计到 2009 年 9 月截止

影响档	很弱	弱	中	强	很强	极强	论文总数
总引用	0—2	3—9	10—29	30—99	100—299	≥300	
与境外作者合作	3669	4244	3563	1271	168	20	12935
中国作者独立	15403	9659	5191	1270	110	7	31640
总计	19072	13903	8754	2541	278	27	44575

年发表的达到了“很强”影响档以上的论文数基本不再增加.换句话说,即使这些达到了“很强”影响档以上的论文,主要也就是在五年以内受到学界关注.离开真正地产生深远影响,名留物理学发展史尚有很大的距离.

表 1 中列出的 2006 年发表的“强”(甚至“中”)影响档以上的论文,明显地少于 2000—2004 年中的每一年.这不足为怪,因为从发表到引用统计截止的时间还较短,只有三年左右.令人感到不足的是从 2000 到 2004 这五年间,虽然论文总数增加达到 60%以上,但“很强”影响档以上的论文数几乎没有增加,增加的论文数明显地只集中在“中”影响档以下.

可以看出,2000 年以后研究生招生的增加,导致初涉研究生涯的年青人大量增长,从培养未来的研究人才上看,倒也不无可取之处.但是从增加的论文大都集中在“中”影响档以下,而“很强”影响档以上的论文并无增加看,可以说中国的物理研究,目前基本上也只是实现了数量和规模的扩张,而要做到质量和贡献的提高,恐怕尚需新的举措和更长期的努力.

3 中国作者的国际合作

除了独立地研究外,最近三十余年来,中国作者与境外作者也有很多合作.表 2 给出了中国大陆作者与境外作者合作,在 2000—2004 年发表的物理论文的引用分布,并与中国大陆作者独立发表的物理论文的引用分布加以比较.

中国大陆作者独立发表的论文总数几乎是与境外作者合作发表论文数的近三倍,这说明大陆有一支很大的研究队伍,能独立地进行物理学研究.而且中国作者不乏独立完成“中”或“强”影响档论文的能力.但是中国作者独立完成“很强”和“极强”影响档论文的能力还不足,“很强”影响档论文的 60%,和“极强”影响档论文的 3/4 还是与境外作者合作完成.可以说,国际合作对提高我国物理论文的影响还

是比较有效的.

我国当前正推行一些措施引进高端人才,如“千人计划”等等.应该把这类人才引进的标准放在哪里?笔者以为在物理学领域里,在可以用研究论文为衡量业绩的工作中,应该要求他们不但自己有产出“极强”影响档论文的能力,而且更重要的是希望他们能在国内建立或带领一个或若干研究团队,提高这些团队从产出“强”影响档的论文到产出“很强”和“极强”影响档的论文的能力.这些引进的人才固然是来去自由,但引进他们后种下的种子要能扎下根,结出果来.

4 两类研究单位的比较

在中国从事物理研究并发表物理论文的单位里,有两类占很大的比重.一类是隶属中国科学院的十多个(物理专业研究)到数十个研究所(物理相关领域),另一类是隶属教育系统的上百(研究型重点大学)到上千所大学.在中国作者独立发表的物理论文中,表 3 列出了中国科学院作者与国内其他单位(主要是大学)作者所发表的论文的引用分布,并加以比较.

中国科学院作者与其他单位独立发表的论文总数之比约为 1:4.7,这不难理解,因为中国科学院所属单位的总人数较少.一般人预期,如果考察“强”或“很强”影响档以上的论文,专业从事研究的中国科学院应该占较显著的比例.因为其他单位是以大学为主,这些单位主要担负着大量培养硕士生和博士生的任务,是在培养学生的过程中产生的研究论文.但在表 3 的数据中可以看到,以绝对数量而论,其他单位独立发表的“强”或“很强”影响档以上的论文数也已达到约 3—4 倍于中国科学院作者的水平.以“强”影响档以上的论文在总论文数中所占的比例而论,其他单位为 4.0%,中国科学院也只有 5.6%,并未见到明显的差别.

表 3 中国科学院作者与中国其他单位作者的论文数,按总引用次数分档排列.

数据来源:Web of Science; 刊物:所有 SCI 物理类(约 260 种); 论文发表时间:2000—2004 年; 总引用次数累计到 2009 年 9 月截止

影响档	很弱	弱	中	强	很强	极强	论文总数
总引用	0—2	3—9	10—29	30—99	100—299	≥ 300	
中国科学院 (无境内其他单位或境外合作)	2314	1337	850	248	18	2	4769
其他境内单位 (无中国科学院或境外合作)	11186	6729	3529	828	70	4	22346
中国科学院 与其他境内单位合作	1903	1593	812	194	22	1	4525
总计	15403	9659	5191	1270	110	7	31640

从一方面看,大学的教师边教书,边带领研究生和自己一起从事研究,即所谓建设研究型大学,这是应该提倡的.从物理学领域的论文产出看,这些年来也的确有了不小的成效.

从另一方面看,中国科学院这样的专业研究单位产出的物理论文,也有与其他单位趋同的态势,近 77% 的大量论文在“弱”影响档以下.就笔者接触,不少中国科学院的研究所也是大量地招收硕士或博士生.在这些本应专业从事研究的单位,同样是处于学习和入门阶段的学生多.一些优秀的研究员也常常在为学生多而不精,不得不把不少精力放在初级人才培养上,甚至要为帮他们做足毕业所需的论文数而伤神.而较有经验的中级人员(如博士后或助理研究员等)严重不足,这样一来,学生过量反而成了提高研究质量的障碍.不单物理界,其他领域也类似,医学生物界的童坦君先生等就撰写了《关于改革博士后制度和壮大博士后队伍的建议》^[1],同样指出我国现在科研一线的主力是硕士生和博士生,而非已训练有素的博士后,研究工作中缺乏较有经验的中级人员.为求改变,童坦君先生等还提出了六项具体建议,笔者以为也同样适用于我国物理界的研究单位.

总之,论文影响分布上表现出了中国科学院与其他单位趋同的态势.是喜是忧,笔者不敢妄断,在此提请有关领导注意和思考.

5 “高端”刊物上的中国论文

我国学术界谈论期刊时,经常贴以“高端”(或“核心”,“优秀”,甚至“顶尖”)等等标签.中国的许多研究单位,乃至国家级的科研管理机构,都把要求或鼓励中国作者在“高端”刊物上发表论文作为引导提高研究水平的一项重要措施.这种圈定刊物发表论

文的作法,大约可以追溯到 1986 年,李政道先生在创建中国高等科学技术中心时就指定了十几种物理刊物作为中心成员发表论文的平台.对中国高等科学技术中心这样一个成员多分散广的松散组合,为保证所发表的论文达到适当的国际交流水准,这不失为一个简洁明智的管理举措.后来从大约上世纪九十年代中期开始,我国一些论文产出已达到较多数量的单位,也开始鼓励成员到一些选稿门坎较高的刊物上去发表论文,这对提高单位的研究水准也曾起过积极作用.

表 4 列出了较受中国作者关注的 13 个物理刊物(刊名见表的下端),以及 2000—2004 年间,在这 13 个刊物上,中国大陆作者独立发表的论文数,仍按引用的强弱分档排列,并与发表在所有 SCI 物理刊物上的论文加以比较.可以看到,中国作者的论文有 1/8 是在这 13 个刊物上发表,已是相当集中;而“强”影响档以上的论文更是有超过 1/3 出现在这 13 个刊物上,“极强”影响档的有 2/3 都集中出现在这 13 个刊物上.

在“高端”刊物中,国内物理界最偏爱的当推 PRL (Physical Review Letters). 一些单位,乃至国家级科技主管部门,都对 PRL 论文青睐有加,甚至是作为优秀成果的象征:每出一篇 PRL 论文,即出简报,写新闻,发奖励;对 PRL 论文的作者,在录用定级,评职考核,审批项目,乃至挑选“杰出”“学者”中也高看一等.如果只按“强”影响档以上的论文在一个刊物里所占的比例看,13 个刊物中 PRL 最高,超过了 40%,远高于其他 12 个刊物.因此,可以说 PRL 的选稿门坎的确较高,上述偏爱也算事出有因.

但这种偏爱显然也有两点误差,甚至造成了两个很大的误区:一是以偏概全,把影响强的比例较高当成 PRL 的论文影响都很强.实际上,中国作者发表在

表4 中国大陆作者独立发表(无境外合作者)在13个刊物上的物理论文的引用强弱分布.

数据来源: Web of Science; 所有刊物指 SCI 物理类约 260 种; 发表时间: 2000—2004 年; 总引用次数累计到 2009 年 9 月截止

引用分档	很弱	弱	中	强	很强	极强	论文数总计
总引用	0—2	3—9	10—29	30—99	100—299	≥300	
发表刊物:PRL	4	18	34	35	6	1	98
发表刊物:APL	51	147	220	143	16	3	580
发表刊物:PR5	295	576	591	199	10	0	1671
发表刊物:JP4	366	396	257	42	3	0	1064
发表刊物:PL2	320	321	259	90	15	2	1007
13 刊物总计	1036	1458	1361	509	50	6	4420
所有 SCI 刊物	15403	9659	5191	1270	110	7	31640

PRL 系指 Physical Review Letters

APL 系指 Applied Physics Letters

PR5 系指 Physical Review A, B, C, D, and E 等 5 种刊物

JP4 系指 Journal of Physics A, B, Cond. Matt, and D 等 4 种刊物

PL2 系指 Physics Letters A, and B 等 2 种刊物

PRL 上的论文中,有近 60%的引用还在“中”影响档以下;二是一叶障目,以为 PRL 外的其他刊物里没有或少有影响强的论文.表中数据给出的结果正好相反:从绝对数看,中国作者的“强”影响档以上的论文出现在其他刊物上的要多得多.中国作者的“强”影响档以上的论文,在 PRL 上发表的只占 3%,而其他刊物占 97%;中国作者的“很强”影响档以上的论文,在 PRL 上发表的也只占 6%,而其他刊物占 94%.所以,以为只有 PRL 里才有影响强的论文,而忽略了其他刊物,会把影响强的论文丢掉 90%以上.

6 依靠专家评价科学论文

看过上面几节表征我国物理论文状况的数据后,笔者还想对我国科学论文评价的现状和其中存在的问题做一些讨论.因为如何评价论文的科学价值,并通过对论文的评价去考核一个作者或一群作者的研究成绩和能力,足以左右大多数科学家,从而左右整个国家的学术发展的效率和方向.

正确地衡量论文的科学价值是一项很困难的工作.笔者以为理想的评价办法只能是依靠专家评议,即表 5 中列在第一位的方法.如果有一定数量的专家,他们专业素养高(专业上水平高,公正地参加评价),又肯付出一定的劳动和时间,经过他们的认真评议应能最准确地判断论文的科学价值.在笔者曾参加的国家自然科学奖评选中,就见过一位专家花了一个多月的时间认真负责地推导和核查论文的每

个公式.这种依靠专家评议的方法,相比于其他基于任何统计数据的评价方法,如论文数量,论文发表刊物的声誉,以及论文被引用的情况等等,都具有无法替代的优势.在评价一些最重大的科学成就时一定是采用这种专家评议的办法,比如诺贝尔奖,以及我国的国家级自然科学奖项的评议等等.

表5 论文评价中常见的四种方法.它们对评审人的专业水准和所需付出的劳动时间的要求,以及可能的延迟时间和评审结果准确性的比较.

评价依据	专业要求	劳动时间	延迟时间	准确性
专家评议	很高	天—月	很短—数年	很高—很低
引用情况	基本	小时—天	数年	高
刊物声誉	无	0	0	中
论文数量	无	0	0	低

其实不但在评价一些最重大的科学成就时,应该采用这种专家评议的办法,就是一些更日常的评议,比如硕士或博士学位论文的评审,青年教授或研究员的聘任升级等等,专家评议也应该是最主要的依据.一个科学家,一个研究单位,一个校系,甚至整个国家范围的科学共同体,它的科学品味高不高,有没有适合科学茁壮生长的文化,笔者认为最重要的体现就是它能不能作好科学论文,以及其他科学业绩的专家评议.

但是,即使是高水平的专家评议,对有些论文的科学价值,在经过反复的实践检验前也还难于做出

结论并取得共识. 这会导致从发表到评价有一段等待或延迟时间. 由于成本较高, 又有延迟, 这种依靠专家评议的方法并不能在各种论文评价的活动中全都采用.

此外, 这种专家评议的方法毕竟又在很大程度上依赖于人, 而且往往是不多几个人, 因此带有较多的主观性, 会有失效的时候: 由于受聘参与的专家不够专业, 或不够公正, 或不够认真, 或时间所限等等, 评价的准确性也可能变成很低. 当评议的专家是由被评议的人或单位自己选定时, 评议常常无异虚设, 结论可能全不可信. 比如, 把假冒的芯片鉴定成了重大成果等也曾经发生^[2]. 从这个意义上看, 反不如其他基于统计数据的评价, 它们较少人为造作的可能, 较少产生出完全不可信的结论.

笔者深感不幸的是, 不被人们信任的专家评议在我国实在不少. 以至于在诟病之余, 人们往往求助于那些较为客观, 较少人为影响的评价方法. 我在表 5 中列出了专家评议, 以及其他几种常见的评价方法, 它们对评审人的专业水准和所需花费的劳动时间的要求, 从论文发表到评价所需的延迟时间, 以及评价结果的准确性, 以兹比较, 并在下面几节予以讨论.

7 依据论文数量评价科学论文

在不算重大成就的评价活动, 比如硕士或博士学位论文由答辩委员会评定, 原本也属于专家评议一类. 但在我国大多数硕士或博士培养单位, 目前更流行的是以发表的论文数量决定是否毕业和授予学位, 比如要求硕士有 1—2 篇, 博士有 2—3 篇论文, 而且基本上都把这个要求放在答辩委员会决定之上, 作为毕业和授予学位的先决条件. 这既含有对答辩委员会(基本由导师本人提名组成, 罕见有答辩通不过的)评定结果的不信任, 或许培养单位也还有基于提高论文产出量的考虑. 也就是说我国在硕士或博士学位的评定中, 实行的基本上是表 5 中所列的第 4 种评价办法, 即依据论文数量评价的方法.

这种按论文数量评价的方法, 对评审人无专业要求, 也几乎不花劳动时间, 且完全没有延迟, 论文一经发表, 评价就有结论, 因此很容易实行. 问题是这种办法几乎不可能反映出论文的科学价值或作者的学术水平. 以此衡量成绩时, 往往还导致不良的倾向: 比如笔者见到过一个组研究 AB 合金的两篇论文, 分成 0—50% 和 50%—100% 两段, 物理内容本

来没有什么不同, 连文字都大同小异, 只是写成两篇以增加论文数量等等.

更有甚者, 就演变出一稿多投或抄袭等不端行为. 固然这类学术不端的主因还应归于作者的道德缺欠, 不能认为是评价方法不当所致, 不过社会学家们在讨论制度与道德的关系时早就有很高明的论述. 他们说, 不能期待人们用道德的完善去填补制度的缺欠, 相反应该用制度的健全去引导人们不断走向道德的完整. 以此衡量, 依据论文数量的评价应该算是有缺欠的制度, 它不能引导人们走向道德完整, 反而还会放大道德的缺欠, 诱发不端的行为.

在 30 年前, 中国刚开始提倡公开发表论文时, 这种按论文数量的评价曾为不少单位所认可. 到现在, 在一些物理研究尚属初创的单位, 增加论文数量也仍是重要的考虑. 表 1 中看到从 2000—2004 年, 中国物理论文的增加几乎都集中在“很弱”和“弱”影响档上, 笔者以为这与我国的论文评价中, 尤其是对硕士或博士学位的评定中, 按论文数量评价仍然盛行不无关系. 是不是那些现在论文数已经不少的单位, 可以率先改变一下思维, 不必在意论文的总数量或总引用次数, 而多看看如何增加“很强”和“极强”影响档的论文. 当然更希望在这些单位之上, 高至国家级的领导管理或资助的单位站得更高, 及早把我国的物理研究从规模扩张引导到质量提高上.

8 依据刊物声誉评价科学论文

前面说过, 从大约上世纪八十年代中和九十年代开始, 我国一些单位就开始鼓励其成员争取到一些“高端”刊物上去发表论文, 对提高各单位的研究水准颇有积极作用. 不过后来不断演变, 导致了当前在我国的科学评价当中盛行一种风气, 即依据发表论文的期刊声誉去评价论文的水平或科学价值, 即表 5 中所列的第 3 种评价方法. 这种评价方法不但在我国的物理界, 而且在我国的自然科学各界, 各单位乃至国家级的科研管理部门中也广为采用.

作者选择哪个刊物去发表论文, 可能出于种种考虑. 比如年轻的学生希望通过论文在“高端”发表, 来尽快提高自己的声誉; 年轻的甚至是已经很有成就的人, 希望通过“高端”刊物的严格审稿, 使自己的工作得以完善; 不过也有很优秀重大的成就, 包括后来得了诺贝尔奖的论文, 作者不愿因“高端”刊物的挑剔耽误时间, 或不愿在发表前被匿名的他人先睹, 反而选择在并非“高端”的刊物发表. 总之, 作者可以

选择他认为最合适的地方去发表,这都可以理解.甚至,有的大型科学设施的投资机构要求该设施上产生的数据或结果一定要在某些指定的刊物发表,这也无可非议.

但是这都不应该影响对论文的科学评价,学界或管理机构或投资机构都不应以发表期刊的声誉如何,是否在自己指定的范围内等等为依据进行评价,任何程度的这类考量都一定是有害无益的.在前面我们已举出了详细的数据,说明用期刊声誉为依据去评价论文的影响可能产生很大的错误:比如,以为PRL论文的影响都强,错误可能达到60%;而以为影响强的论文都在PRL里,错误会超过90%.评价论文的影响力尚且如此,不用说以期刊声誉为依据去评价论文的科学价值就更不可信了.

在当前我国颇为流行的科学评价中,更加近乎荒谬的是把期刊的声誉又简化成用影响因子去衡量.对高影响因子的期刊达到几乎崇拜的地步,比如有单位把硕士或博士毕业的要求定为发表论文的期刊的影响因子之和达到一定指标,也有单位在教授聘任或升级时要求其发表论文期刊的影响因子之和达到多少等等.殊不知所谓影响因子只反映一个刊物在前两年发表的论文的平均被引用的多少.因为同一期刊的不同论文的引用差异百倍以上,平均值很难代表每一篇论文的影响;又因为只是两年的引用,并不反映论文的长期(即包括三年及以上)的总引用,所以也不能反映论文的总影响.于是有的以发表热门论文为主,影响因子很高(达到6—9),但引用的半衰期不长(短于3年),每篇论文的平均总引用并不多的刊物,也受到不少国人的热捧.崇拜影响因子,在研究选题上助长了追热门的风气.笔者曾指出^[3],当下我国的物理论文存在一种浮躁现象,即“短平快产品很多:短者指把一篇内容扯成两三篇写;平者指没有突出的创新,快者指昙花一现、寿命不长”,这与我国科研评价中长期流行按期刊影响因子评价不无关系.

论文“短平快”浮躁现象的存在或许也是我国物理研究发展中不得不面对的一个阶段.但至少物理界要正视这个现象的消极面,不要陶醉于论文数的增加,不要崇拜影响因子,也不要崇拜“高端”刊物,从而尽快地走过这个阶段.各研究单位的领导,尤其是国家级的科研主管部门,比如科技部、国家基金委员会、教育部、科学院等等更责无旁贷.检讨和改进论文的评价方法,应该是一个重要的方面.

9 依据引用情况评价科学论文

在感到上述不足之处后,表5中所列的第2种评价方法,即依据引用情况评价论文的方法,已经开始受到关注,并在一些单位开始采用.这种方法根据论文在发表后一段时间里被引用的情况(比如总的引用数量,他人引用的数量,引用人的地域或工作单位的广泛性,引用随时间增减的快慢,以及引用中给与的评价等等),去考察论文在科学知识积累过程中所起的作用.收集和鉴定这些数据,对评价人的专业有要求,但并不很高,所需付出的劳动时间也不太长,因此也能实行.原始的引用普遍来源于学术界,少数人的爱恶难于改变评价结论,又均系公开发表,若有猫腻亦当有迹可寻,而且是通过了一段较长时间的考验,因此准确性比较高.

比如,要考察我国在2000—2004年间发表的近3万多篇物理论文中那些产生了强烈影响,只要花1—2小时就可以从数据库中检出如表1所列的数据.当然,科学价值的高低不完全等同于影响的大小,影响的大小更不能只用引用的多少去衡量,不同的学科分支又有很大的差异.因此还必须再按学科分支请几位专家,每人花1—2天时间,针对“很强”影响档以上的约110篇论文,鉴别引用的具体内容,考察引用中给与的评价,这样一定可以找出影响力真强的论文.显然,这样做的准确性远比只看发表刊物的声誉为高.

这种基于引用情况的评价,从发表到评价毕竟还有一段等待或延迟时间,因此在一些时间性很强的评价中,例如要在硕士生或博士生毕业前对他们的论文做出评价时无法采用.这种基于引用情况的评价方法在其他许多评价活动中,比如检查一个研究组,一个实验室,一个研究所的研究业绩时,是完全可以采用的.因为在这样的评价中,完全可以,而且也应该不拘泥于1—2年,甚至不拘泥于3—5年,而是从更长的时间尺度上去考量.

时间尺度多长为宜?笔者建议可以参考物理期刊的引用半衰期的长短.除去引用半衰期很长的综述性期刊外,所有SCI物理刊物的引用半衰期平均约为5—6年(各年的数据略有涨落).因此衡量物理论文的影响,用6—10年作时间尺度应该才是合理的.当然,6—10年也只是个平均数,对于有的学科分支也算不上足够长久,一篇优秀论文也可能还未引起学术界重视,引用不多,或理论尚未被实验证

实,或实验尚受到质疑而未获肯定等等.比如常规超导体的BCS理论,就经过了十多年的质疑和争论才最终被肯定,获得了诺贝尔奖.因此,6—10年应该认为只是一个大体合理,尚可的操作的时间尺度,而不能把一定要在6—10年里产生多少影响作为评价论文科学价值的固定标准.

现在我国的各级领导在同一岗位上的时间都不会太长,用6—10年的时间尺度去衡量,难以表现当届的政绩,不知道他们会不会乐于采用?当然,用6—10年的时间尺度,看看上届或上上届的成绩并加以继承发扬,把对自己的评说留给下届,既得宁静又可致远,似也不失开明.

此外,要依据引用情况评价论文就一定要撇去论文是否在“高端”刊物上发表等等考量,完全根据论文产生的影响,实事求是地进行评价.不问何处发表,那怕不在任何刊物发表,而只在网站上公开,只要在科学界产生了足够的影响,就应该对这些论文完全平等地给予积极的评价.从对参与评价的人或管理部门提出的要求看,这样做比只看论文数量或只看刊物声誉(影响因子)要求要高.不过在管理部门工作的人,现在也大都是硕士博士,再聘请一些专家参与和协助,应该有条件做到.

关键还是在管理层的思维上要有一些改变.笔者想举一个最近发生在我国最高级别的科研管理部门的反面事例,说明思维上存在的问题.这个事例发生在科技部基础研究管理中心发表的“2008年度中国基础研究十大新闻简介”中^[4].由于我国科学家去年在铁基超导研究上做出了突出的贡献,这一系列研究工作被选入十大新闻.“简介”中列举出中国科学家“报道了转变温度为43K的SmFeAsO_F和41K的CeOFeAs,“又将超导转变温度进一步提升到55K,这是目前铁基超导材料的最高转变温度”,和“首个报道空穴掺杂型铁基超导材料”等主要成就,并介绍说“上述研究进展已发表在Nature和Europhysics Letters等期刊上,在国际上引起了巨大反响”.表6列出了“简介”中提到的这4篇铁基超导论文的发表及引用数据.

这4篇论文,在发表后仅半年(到2008年底)就被引用了79到112次,其中国外机构引用39到66次;经过一年多(到今年9月底),总引用增加到201到380次,其中国外机构引用增加到94到218次.的确是“在国际上引起了巨大反响”,列入十大新闻当之无愧.但在这几篇之间比较,无论是在2008年底评选十大新闻时,还是到最近(2009年9月),也

无论是总引用还是反映国际反响的境外引用,发表在国内英文版期刊“中国物理快报”(Chin. Phys. Lett.)上的那篇论文都高于“欧洲物理快报”(EPL).而科技部基础研究管理中心在列举“国际上引起了巨大反响”时,不知道出于何种原因,显然忽略了发表在国内刊物上产生的同样巨大的反响.倘若科技部基础研究管理中心,在评选或发布新闻前真看过表6中的数据,笔者以为他们一定不会有歧视国内的刊物.笔者猜想,多半只是下意识地觉得国内刊物上的论文不大可能产生多大的国际影响,也就是前面第5节中说到的“一叶障目”的错误.而这样低俗的错误竟犯在国家级的科研管理部门,使笔者感到要做到完全根据论文产生的影响,实事求是地评价论文,从我们最高级的科研管理部门开始,在我国科学界的思维上真应该做些改变.

10 结束语

科学论文的作用不单在于它是学术交流的平台,而且更加重要是,通过论文内容的引用和证辩(证实或证伪),实现科学知识的积累,这使公开发表论文成为不可替代的,最重要的知识积累的载体.从本文列举的数据可以看到,我国物理论文的数量在2000年以来一直有很快的增长.但从论文的引用考察我国的物理论文,考察它们在科学知识积累过程中所起的作用和影响时,值得注意和思考的是,从2000—2004五年间发表的论文的引用看,具有很强影响力的论文数几乎没有增加.

从研究人才上看,中国大陆作者独立完成“中”或“强”影响档论文的能力虽然不低,但独立完成“很强”和“极强”影响档论文的能力还有欠缺.这与研究论文绝大多数都是在硕士生和博士生的培养过程中完成,而博士后和其他已有相当研究经验的中级研究人员在国内各单位中普遍短缺有关.值得期待的是,通过国际合作和人才引进,通过改革博士后制度和壮大博士后队伍,能在国内促进一些研究团队,提升产出“很强”和“极强”影响档的论文.

中国的许多研究单位,乃至国家级的科研管理机构,过去把要求或鼓励中国作者在一些“高端”刊物上发表论文作为引导提高研究水平的一项重要措施,有一定道理.但是,把是否发表在“高端”刊物作为评价论文的影响力,并进而作为评价其科学价值的依据,则错误很大.正确评价科学论文的影响,应该是不问何处发表,完全根据论文产生的实际影响,

表 6 四篇铁基超导理论论文的发表及引用. 境外引用是指引用论文的作者全部为境外机构. 总引频次和境外引用后的两个数字分别为 2008 年和 2009 年(到 2009 年 9 月底为止)的引用次数

标题: Superconductivity at 43 K in SmFeAsOF 作者: Chen X H, Wu T, Wu G <i>et al.</i>	刊物: Nature, JUN 2008 卷: 453, 期:7196 页:761 总引频次: 112+268 境外引用: 66+152
标题: Superconductivity at 55K in iron-based F-doped layered quaternary compound Sm[OF]FeAs 作者: Ren Z A, Lu W, Yang J <i>et al.</i>	刊物: CHIN. PHYS. LETT. JUN 2008 卷:25 期:6 页:2215 总引频次: 96+215 境外引用: 51+140
标题: Superconductivity at 25K in hole -doped LaSrOFeAs 作者: Wen H H, Mu G, Fang L <i>et al.</i>	刊物: EPL, APR 2008 卷:82 期:1 文献编号:17009 总引频次: 86+115 境外引用: 41+53
标题: Superconductivity at 41 K and its competition with spin density wave instability in layered CeOFFeAs 作者: Chen G F, Li Z, Wu D <i>et al.</i>	刊物: PHYS. REV. LETT. JUN 2008 卷:100 期:24 文献编号:247002 总引频次: 79+218 境外引用: 39+115

平等地和实事求是地给予评价. 衡量物理论文的影响, 时间尺度不应太短, 至少要 6—10 年以上为宜. 定量的评价以引用情况为依据, 可能有助于促进我国的物理研究尽快走过“短平快”的浮躁阶段, 形成深入严谨的学术研究氛围.

不过归根到底, 每位教授的科学品味高不高, 每个研究单位, 甚至整个国家范围的科学共同体有没有正确的价值取向, 才是最根本最重要的. 这个科学上的品味和价值取向有一个最重要的体现, 就是它能不能自己作好专家评议, 包括科学论文以及其他科学业绩, 从学位授予, 到课题审批, 再到国家奖励等等. 这是任何外来的帮助都无法取代的, 国外评审也好, 引用统计也好, 可以是很好的参考和辅助, 但再好也只是参考和辅助. 科学研究在本质上是一个民主的社会活动, 只有科学共同体的正确价值取向才能保证它的创新能力.

2008 年在中国科学院物理研究所成立八十周年的纪念会上, 杨振宁先生提及, 100 年前当爱因斯坦等开创 20 世纪物理学的重大突破时, 中国不但读得懂的人没有, 甚至连关心和注意人都没有. 笔者记得也曾读过《1956—1967 年科学技术发展远景规划纲要(草案)》, 其中对当年我国物理学人才状况的描述说到, 那时全中国境内曾经发表过物理论文的人

不过 200 左右. 抚今追昔, 早已是今非昔比: 每年产出的达到“强”影响档级别的论文就有 200 多篇. 笔者深信, 假以时日, 我国物理界必能走进世界的前列, 对科学的发展做出更大的贡献.

后记 过去两年里, 在一些讨论和会议中, 笔者向物理界和科技界同仁报告过本文中的数据 and 想法. 承蒙他们不吝指教, 使不足和错误之处得以改正. 特此致谢, 恕不列名.

参考文献

- [1] 童坦君等.《关于改革博士后制度和壮大博士后队伍的建议》, 2009 年 3 月, 中国科学院文件
- [2] 原上海交通大学教授, 微电子学院院长陈进, 在 2003 年称研制成功“汉芯”芯片, 并骗过了鉴定专家、大学当局、地方政府和中央有关部委, 曾名噪一时(比如, 新华社 2003 年 2 月 27 日稿“国内首个高性能 DSP 芯片‘汉芯一号’研制成功”, 或《中国教育报》2003 年 3 月 12 日报道“走近‘中国芯’——访上海交大芯片与系统研究中心”等等). 后经揭发纯属造假, 责任人也受到相应处分(见上海交通大学 2006 年 5 月 12 日所发“关于‘汉芯’系列芯片涉嫌造假的调查结论与处理意见的通报”)
- [3] “评价科研论文莫入误区”, 科技日报(2008—6—25)发表的对笔者的采访记录之一
- [4] “2008 年度中国基础研究十大新闻简介”(科技部基础研究管理中心), 载于“中国基础科学”11 卷总第 67 期, 2009 年 1 月