



上海市学术著作出版基金

徐剑 著

中国网络版权侵权实证研究

基于 3004 份司法裁判文书的量化分析



 上海人民出版社



上海市学术著作出版基金

徐剑 著

中国网络版权侵权实证研究

基于 3004 份司法裁判文书的量化分析

 上海人民出版社

图书在版编目(CIP)数据

中国网络版权侵权实证研究:基于 3004 份司法裁判文书的量化分析/徐剑著.—上海:上海人民出版社, 2017

ISBN 978-7-208-14571-9

I. ①中… II. ①徐… III. ①互联网络-版权-侵权行为-研究-中国 IV. ①D923.414

中国版本图书馆 CIP 数据核字(2017)第 155064 号

责任编辑 肖 峰

中国网络版权侵权实证研究

——基于 3004 份司法裁判文书的量化分析

徐剑 著

世纪出版集团

上海人民出版社出版

(200001 上海福建中路 193 号 www.ewen.co)

世纪出版集团发行中心发行 上海商务联西印刷有限公司印刷

开本 635×965 1/16 印张 15.5 插页 5 字数 180,000

2017 年 8 月第 1 版 2017 年 8 月第 1 次印刷

ISBN 978-7-208-14571-9/D·3050

定价 48.00 元

中国网络版权侵权实证研究
基于 3004 份司法裁判文书的量化分析



抑制理性的有限性——代序

徐剑教授的研究成果即要付梓成书，他嘱我写篇序言。十多年来，我和徐剑是同事、师生和朋友，我有责任也很高兴做这件事。

徐剑研究的是知识产权问题，更具体些说，是网络知识产权的司法裁判问题。他主要回答了两个问题：一是我国网络知识产权案件的裁决中是否存在地方保护；二是在涉外网络知识产权的案件中，境外原告是否享受了“超国民待遇”。对于知识产权，虽然这些年来我从徐剑那里了解到一些相关的知识，但仍然只能算是外行人，讲不出有见地的看法。我想要评论的是徐剑采用的方法，因为多年来我讲授社会研究方法方面的课程，对此有些心得，也因为徐剑的研究方法在我国法学研究中不常见到，可为思考方法问题提供新鲜的思路。

按照徐剑的概括，本书的方法特点有三个，一是数据全面——他称为“全景式”的；二是数量化分析，具体些说，是使用了传播学中量化文本内容的分析方法；三是运用经济学的理性人假设，从利益最大化的角度解释主体的行为动机和策略。绝大部分知识产权的官司是经济官司，侵权为了钱，维权也主要为了钱，用理性人假设来解释打知识产权官司的行为，合情合理。我打算讨论的是前两个特点，数据采集和量化分析，评价它们在本书中的作用。

在《有限理性模型及其他经济学论题》(Models of Bounded Rationality and Other Topics in Economics, 1982)一书中,赫伯特·西蒙(司马贺)解释过他那个著名的有限理性模型。在西蒙看来,任何决策者都会遇到三个约束:其一,就有哪些可能的选择和不同选择可能带来的后果而言,人们只能得到和使用十分有限的信息,且通常不那么可靠;其二,就评价或加工信息而言,人脑的能力有限,会出现偏差;其三,就决策时间而言,人们通常需要在有限的时间内作出决策或判断。因此,在复杂情境下,人们能够作出的决策或判断通常是趋向于满意的(satisficing),而非是最优或利益最大化的。这位图灵奖和诺贝尔经济学奖的获得者断言,理性的这些有限性使人们几乎无法考虑到每一种可能的偶发事件,在更多的时候,决策与判断不得不依赖于所谓“经验法则”(rules of thumb),类似于“差不多”,“八九不离十”,或“经验之谈”。

具体到社会科学研究,由于关涉人类、关涉社会的信息浩如烟海,我们所能获得的信息通常是沧海一粟。而且,人类的主动性、人的心理、观念、行为、组织都可能变动不居。今天获得的信息,明天也许就变化了,关于这群人的信息,用到那群人可能会不靠谱,获得稳定可靠的信息并不容易。西蒙模式中的第一种约束,即信息的有限性,在研究与人、与社会有关的问题时体现得十分明显。又由于受到演化过程、群体认同、宗教、意识形态、文化风习、政治立场和物质利益的影响,如卡尼曼和特沃斯基发现的种种认知偏差,社会心理学家、社会学家所发现的形形色色的社会偏见,无处不在地制约乃至瓦解人们准确处理信息的能力。西蒙模式中的第二种约束,即信息加工能力的有限性,在社会认知研究中也体现得淋漓尽致。对此卡尼曼和特沃斯基有过一系列实验研究。从学数十年,我接触过许多关于人和社会的发现、观点和理论,深知判断哪些是真知灼见,哪些是

误解,哪些是不着边际乃至居心叵测的误导,实在是非常困难。

如果说,有限理性是人类认知的宿命,但同时也假定学术研究以尽可能接近真实为目标,怎样才能抑制理性有限性对我们的影响?

作为司法裁判研究的成果,徐剑的研究占有三点优势或运气。首先,宏观地看时代背景,借助计算机和网络技术的发展,在社会科学研究不少领域,研究者有了获得大量乃至全部相关数据的可能,因此也有了大数据研究的势头。技术演进带来的这一可能性会改善信息有限性的约束。其次,从具体的社会条件去看,2006年3月10日,中国知识产权裁判文书网(<http://ipr.chinacourt.org/>)开通,全面公开全国法院知识产权类裁判文书,这使徐剑能够采集到2002年到2011年十年间网络知识产权的3004份法律裁判文书(原则上,这当是该时间段中网络知识产权司法裁判文书的总体)。第三,从具体研究对象去看,与其他文献材料相比,例如历史典籍、文学作品或新闻报道,司法裁判文书是结构化程度较高的文本,原告、被告、裁判地点、审判员、代理人、裁判时长、裁判结果等研究所需的关键指标,均可以客观且较容易地从文本中提取,并加以数量化。在我看来,这三点优势降低了理性有限性对研究的影响,提高了徐剑发现的说服力。

以徐剑对第一个问题的回答为例。基于地方性数据和个别案例分析,过往研究大多主张,中国司法裁判存在一定的地方保护主义。这样的看法很容易得到我——也许还有更多人——的赞同。虽然我明白自己对网络知识产权案件知之甚少,但地方保护的结论却符合我的一般经验。比如,我知道在中国的体制中,由于政绩考核要求,政府会有保护本地企业,增加本地收入的动机,我还知道在中国社会中,血缘—地缘是形成社会网络的基础,而这种网络很可能对司法裁决发生影响。换言之,西蒙所谓“经验法则”很可能支配我的判断,尽管我参考的这些信息与所要回答的问题未必有实质性联系。

徐剑在研究中发现,就确认侵权而言,“和以往研究的预期相反,异地原告胜诉率更高”,“饱受抨击的地方保护主义倾向在网络版权侵权案件的判决中并未明显存在”。由于使用了总体数据,避免了用局部样本表示总体可能发生的以点带面的片面性,徐剑的发现应更接近事实真相,会更具说服力。资料全面还常意味着可分析的数据分类增多,允许徐剑做更深入的探索。当对原告、被告和法院的属地,原告、被告的自然人或法人身份,原告、被告是否为政府机构、事业单位作出细致的多重区分之后,徐剑进一步发现,法院对于本地原告的经济赔偿诉求支持力度更大,体现为裁决赔偿额在原告要求赔偿额中所占比例更高,对本地政府部门,事业单位被告的经济赔偿裁判更为温和,裁决赔偿额在原告要求赔偿额中所占比例低于平均水平。按照徐剑的话来说,在网络知识产权的司法裁判中,地方保护通过“隐晦”的方式来实现,远非如一般直觉到的那样普遍和明显。更有意义的是,细致的数量化分析有助于抑制“经验原则”在判断中的影响,从而改善了认知能力带来的偏差。比如,这些数据分析抑制了联想的空间,让我很难再把血缘—地缘型人际网络理论当作启发性线索(heuristic cue)或有据推测(educated guess)的根据,对司法裁决的地方保护问题作出过度简单化的判断。

再看徐剑对第二个问题的回答。若问到在网络知识产权案件中,境外权利人是否会得到“超国民待遇”,我,可能还有许多人,多半会给出肯定性回答。原因何在?我这样年纪的人在“文革”时期长大。那时,一方面,我们被告知国外人民生活在水深火热之中,另一方面,我们见到在中国的外国人似乎享受着“超国民待遇”:只有他们坐得起当时看起来价格不菲的出租车;只有他们能够持有外币换来的外汇券,到国人勿入的外汇商店买好东西;他们居住的高级旅馆也称作“涉外单位”,顾名思义,国人通常是不该进去的。即便改革开放

后,外国人来得多了。在大学校园里,我们看到外国教师、留学生有专门的宿舍楼,条件远好于中国老师、学生居住的地方。从新闻报道中,我们得知各级政府给了外国企业不少优惠政策,官员们高接远迎外国客人,也得知不少外国人受到侵害后,比如财物被盗,警方神速破案的故事。虽然我并不真的了解中国法院今天如何对待外国原告和被告,但当年的记忆在大脑里想必是储存了很多。每遇到中国如何对待外国人这类问题,这些记忆便倏然出现,让我更可能相信境外权利人享受到“超国民待遇”。在有限理性的模式里,这样的判断与信息有限的约束有关——我对于中国法院如何对待涉外案件知之甚少,也与评价和加工信息能力的约束有关,如卡尼曼和特沃斯基所谓的“易得性偏见”(availability bias),即记忆中已有的、个人化的、强固记忆的信息最可能影响乃至决定判断结果,尽管这些信息与判断对象没有实质联系。

徐剑的数据告诉我,事实要比易得性偏见带来的判断更为复杂。虽然在网络知识产权的裁决中,境外原告的胜诉率明显超过了境内原告,境外原告也获得了比境内原告更多的赔偿金,这似乎支持“超国民待遇”的看法,但若比较境内与境外原告的诉讼成本(如裁决时间,律师费用等投入)和实际收益(如诉求赔偿金额与实际赔偿之比,成本投入额与实际赔偿额之比),可知境外原告的诉讼成本明显高于境内原告,境外原告所获赔偿额与要求的赔偿额之比明显低于境内原告。虽然我们难以知道,境外原告的高胜诉率究竟是诉讼投入较高所致,还是因为法官对他们另眼相待,予之以“超国民待遇”,我们能够知道的是,若假定境外原告的诉讼动机符合经济理性人假设,高投入与低收益的裁判结果可能让他们意识到打官司其实得不偿失。近年来,在网络知识产权案件中涉外案件占比锐减,似可用来间接证明境外原告并不那么满意他们受到的待遇。

面对这种基于总体数据的证明方式,储藏在我大脑里的那些儿时记忆,那些道听途说的故事便难以进入判断过程,更难以在判断时发生重要影响。从西蒙的模型去看,这是说,信息有限性的改善带来了认知能力的提高,二者正相关。

使用裁判文书总体并将之数量化的研究方法有令人鼓舞的前景,因为采用这样的方法可能描述法学研究中某些领域的全面情况。美国法学家桑斯坦曾把法学研究分为三类:(1)实证(positive)研究,关注对实际情况观察和分析;(2)规则(prescriptive)研究,关注法律怎样实现社会目标;(3)规范(normative)研究,关注法律理应为社会做什么。在比较了法学自然科学和社会科学的方法后,卢宾指出,法学研究传统通常不重视描述(descriptive)或实证研究,规则研究才是重心所在。然而,在逻辑上,实证研究提供对行为实然状态的理解,规范研究建立应然的理想目标,规则研究寻找从实然到应然的路径。缺少对真实状态的把握,规范理论可能只是空中楼阁,规则设计也会沦为无的放矢。准确描述实然状态因此是法学研究者追求的基本目标之一。在观察实然世界时,如认知心理学所揭示的那样,人们难免会为认知约束所羁束。如徐剑的研究所示,越来越多、越来越透明的司法信息可明显改善这些认知约束。

除开徐剑已经做到的,本书体现出的方法进路还有相当大的延伸空间。比如,使用回归分析等方法有可能进一步发现裁判文书中分析出来的一些变量各自对裁判结果的影响程度。当知道了不同变量与历史上原告或被告获胜的概率关系后,有可能使用预测模型,如贝叶斯定律,来预测未知案件的胜率。当了解了胜率在总体中的分布之后,应不难发现每个法院、每位审判员、每家律所、每位律师在案件中的表现。这不但可以为当事人提供有价值的司法信息,也可以提示管理者关注那些远离平均概率的异常值,探求背后的原因是什

么,是案件的特殊性,是有关人员的不寻常能力,还是违背法律的利益交换行为?不断积累起来的大量数据可为机器学习提供素材,最终使人工智能分析和预测司法裁判结果成为可能。若干年前,在对媒体名誉侵权裁决的一项研究中,徐剑便和我仔细讨论过这样的数据方法所蕴含的巨大学术、管理和市场价值。这两年,国外和中国都已经有了商用的法律机器人。我猜想,指导这些机器人工作的应该正是上述原理。

当然,仅仅依赖数据,许多问题难以得到很好的回答,例如,本书发现的“隐晦”的地方保护主义问题。通过数据,我们了解到法院对本地原告的赔偿诉求更加支持,对本地政府部门、事业单位被告的赔偿裁判更为温和,但我们无法知道,法院的行为是由审判人员与当地原告、被告直接或间接的人情关系,或当地政府对司法系统或明或暗的指示,或审判人员与当事人之间的违法交易,或仅是地域认同感带来的群体偏见所致。换言之,数据显示出司法行为是怎样的,却没能告诉我们产生这些行为的社会、制度或心理机制。这样的问题,还需要通过观察、访谈、调查等传统质化方法来解答。

还必须看到,在不少的社会研究领域,全面获得资料或数据是不可能的。例如,我曾学习过考古学。学习中我意识到,考古发现大多有偶然性,难以知道偶然发现的工具、器皿、遗址与当时的社会生活、文化进程整体之间的关系。因此,考古学家不得不把发现与其他间接材料,如其他地区、时代的发现物,人类学家、民族学家对当代原始部落的研究结果,以及理论想象拼合起来,构造出远古生活的图景。这种图景只要在逻辑上合理,且与已知证据不冲突,便可以被接受。如若深究,由于远古物质和精神生活资料总体几乎总是未知的,偶然获得的证据能在多大程度上代表总体几乎总是不确定的,基于偶然和局部资料所构建的远古图景也就难逃不确定性。在这种意义上,

努力追求真相的研究仍会如西蒙所说的那样,只能趋于满意,难以最优。

我的两位同事,认知科学家西蒙的学生秦裕林教授和法学家林喜芬教授对此文的写作提出了很好的建议。我在此感谢他们。

葛 岩

于上海一闵行

2017年5月5日

目 录

抑制理性的有限性——代序	1
绪论	1
第一章 中国版权保护和侵权治理：文献综述与研究设计	10
第一节 中国版权研究综述	10
第二节 国内外盗版研究综述	26
第三节 中国知识产权司法保护研究综述	32
第四节 研究方法和设计	39
第五节 本章总结	48
第二章 中国网络版权侵权的司法实证分析	50
第一节 中国网络版权侵权案件概貌	50
第二节 中国网络版权侵权诉讼的胜诉因素分析	71
第三节 网络版权侵权诉讼的司法效益分析	85
第四节 本章总结	92
第三章 网络版权侵权诉讼中的地方司法保护实证分析	94
第一节 问题的提出	94

第二节	影响变量	97
第三节	原告胜诉率的地域分布	98
第四节	原、被告和法院的地域关系对判决结果的影响	100
第五节	索赔金额对判决结果的影响	102
第六节	原、被告和法院的地域关系对经济赔偿金额的影响	104
第七节	被告身份特征对判决结果的影响	105
第八节	法院审理阶段对判决结果的影响	107
第九节	本章总结	109
第四章	涉外网络版权侵权诉讼的实证分析	111
第一节	问题的提出	111
第二节	涉外网络版权侵权案的概貌	113
第三节	涉外诉讼的时间成本分析	119
第四节	涉外诉讼的被告抗辩投入程度	121
第五节	境外法人原告的投入程度	122
第六节	境外法人原告的预期收益和实际收益	123
第七节	境外法人的案件受理费成本	125
第八节	境内外法人原告直接经济收益的比较	126
第九节	外企和港澳台企业的诉讼行为差异	127
第十节	本章总结	128
第五章	结论与探讨	131
第一节	研究结论	131
第二节	进一步的讨论	135
第三节	研究的贡献	136
第四节	研究的局限及展望	138

附录一 中国媒体名誉侵权司法裁判的实证分析	140
附录二 典型案例	155
附录三 相关法律法规	167
参考文献	214
跋：一种数据型的社会科学观察方法	229

图 表 目 录

图一 全书的研究结构图	8
表 1-1 中国版权研究 50 篇高被引文献列表	11
图 1-1 中国版权研究高被引文献的时间分布	14
表 1-2 高被引文献所属机构分布	15
表 1-3 高被引文献来源期刊分布	16
表 1-4 高被引文献资助基金分布	16
表 1-5 高被引文献关键词分类	17
表 1-6 高被引文献研究类型分类	18
图 1-2 研究路线图	45
图 2-1 中国网络版权侵权案件历年增长图	51
图 2-2 中国网络版权侵权案件历年增长修正图	51
表 2-1 网络版权侵权裁判文书地域分布	52
表 2-2 时间分类和发案地分类的交叉制表	53
图 2-3 北京、上海和其他地区在 2007 年前后的发案数比较	54
表 2-3 网络版权侵权案件的判决形式分布	54
表 2-4 北京、上海和其他地区在诉讼裁判方式上的比较	55
表 2-5 司法诉讼寻求救济的著作权权利类型	55
表 2-6 侵权权利类型在 2007 年前后的发案比较	56
表 2-7 侵权作品类型分布	56

表 2-8	北京、上海和其他地区在侵权作品类型上的比较	57
表 2-9	案发地分类—时间分类和侵权作品类型的交叉分析	58
表 2-10	原告身份类型在 2007 年前后的发案比较	58
表 2-11	原告身份类型和侵权作品的交叉分析	59
表 2-12	北京、上海和其他地区在原告身份类型的比较	60
表 2-13	侵权作品类型和诉讼权利纠纷类型的交叉分析	61
表 2-14	原告身份特征和诉讼裁判方式的交叉分析	61
表 2-15	纠纷权利类型和诉讼裁判方式的交叉分析	62
表 2-16	被告身份类型在 2007 年前后的发案比较	63
表 2-17	北京、上海和其他地区在被告身份类型的比较	63
表 2-18	被告身份类型和时间分类、地域分类的交叉分析	64
表 2-19	被告身份类型和纠纷权利类型的交叉分析	64
表 2-20	被告身份类型和诉讼裁判方式的交叉分析	65
表 2-21	判决书的审理时长分布	66
表 2-22	2007 年前后超时审理案件的比较	66
表 2-23	北京、上海和其他地区在超时审理案件的比较	67
表 2-24	时间分类、地域分类和案件审理时长分类的交叉 分析	67
表 2-25	时间分类、地域分类和案件审理阶段的交叉分析	68
表 2-26	原告的索赔请求明细	68
表 2-27	个人原告和法人原告在收集证据上的比较	69
表 2-28	个人被告和法人被告在收集证据上的比较	70
图 2-4	原告胜诉率的时间分布	72
表 2-29	时间分类和判决结果的交叉分析	72
表 2-30	地域分类和判决结果的交叉分析	73
表 2-31	原告身份特征和判决结果的交叉分析	74