

# 论刑法教义学与实证研究

白建军\*

---

**内容提要：**定性研究与定量分析、规范学与事实学的不同，只是刑法教义学与刑法实证研究之间表面上的不同，其掩盖了两者内在相和的价值。法教义学的实定法合理性假定和法律实证研究的集体经验合理性假定，都是相对的，二者之间不存在根本对立，采用实证方法研究刑法教义学问题并非不可能。刑法教义学与实证研究统一于实践理性，相遇在刑法现象大样本中，交集于法的有效性。没有教义学的刑法实证研究，或者没有实证研究的刑法教义学，都无益于中国刑法研究的全面发展。最极端的恶是非法剥夺他人生命的犯罪，最极端的社会否定性评价是依法剥夺犯罪人生命的死刑，这两个剥夺是其他一切剥夺的边界。以近8万个致命案死刑适用样本的刑罚教义学研究为例说明，刑法教义学实证研究是一种基于刑法现象大样本的描述性批判，其包括刑法理论学说的教义学实证研究、刑法规定的教义学实证研究和刑事司法实践的教义学实证研究。

**关键词：**刑法教义学 实证研究 死刑适用 法的有效性

---

法学研究的学术分化中，〔1〕刑法教义学与刑法实证研究之间既无交锋争讼，也无明显的相互依存。不过，此种“不合作”状态对刑法研究未必有益。有趣的问题是：应怎样理解这两种研究方法之间的关系；能否用实证方法研究刑法教义学问题；如何实现法教义学与实证研究在中国刑法研究中的深度合作。本研究的结论是，刑法教义学与刑法实证研究和而不同。二者显而易见的“不同”中隐含着“和”的条件，“和”才是二者关系中一直以来被忽视的价值。刑法教义学实证研究，就是刑法教义学与实证研究之间和而不同的产物，也暗合储槐植所倡导的刑事一体化思想。

## 一、法教义学与实证研究相和的难点及其化解

一般来说，法教义学多采用演绎逻辑，进行抽象、思辨，而实证研究多采用归纳逻辑，进

---

\* 北京大学法学院教授。

〔1〕 参见谢海定：《法学研究进路的分化与合作》，《法商研究》2014年第5期，第87页。

行观察、描述；法教义学强调典型情形的定性研究，而实证研究强调大样本量化分析；法教义学拥有一套体系性知识，以其系统的自给自足和逻辑的严谨体现其科学性，而开放的实证方法本身并无系统的实体理论，以其“假设—检验”的程序性范式保障其研究结论的可靠性。法教义学与实证研究之间表面上的差异并非它们进行合作的主要障碍，相反，恰恰是实现互补的基础条件。真正妨碍法教义学与实证研究走到一起的，是它们之间的两层隔膜：一是各有各的合理性假定，二是各有各的技术短板。

### （一）实定法合理性假定与集体经验合理性假定

法教义学假定实定法的合理性，而实证研究假定法律实践集体经验的合理性。按照前者，法律不容批判，只能对其做合理解释；按照后者，法律实践的集体理性必有其道理，是公正检验的标准。这就形成了二者之间的紧张关系：既然法律不容批判，就无需其他检验标准；既然实践的集体理性是公正检验的标准，就无需另寻合理性阐释。

法教义学对实定法秩序抱有坚定信仰。即所谓“法教义学要以对一国现行实在法秩序保持确定的信奉为基本前提，这也是所谓的‘教义’的核心要义所在。”〔2〕拉德布鲁赫认为，法教义学是关于实在法的科学，而非应有法的科学。〔3〕法教义学“是关于有效的法的科学，而不是关于正确的法的科学，是关于存在的法的科学，而不是关于应该存在的法的科学”。〔4〕总之，“无法律，不教义”，〔5〕法教义学“假定现行法秩序大体看来是合理的”。〔6〕追问实定法背后的应然与否，就可能走向对实定法的批判，而法律不容批判。

然而，实证研究的一个基本特征是证否。如果一个命题不可证否，就不适宜对其做实证研究。表面上，这与法教义学的基本假定格格不入。不过，且不说实证研究证否的对象更多只是法学理论、学说，就算是法律，也需澄清法教义学所说的实定法到底是指法条本身，还是法律规范。法条是法律规范的载体，而法律规范是隐含在数个法条之间的逻辑关系。这意味着，如果某个具体法条与体系性的法律规范发生冲突，法教义学会毫不犹豫地遵从整体的法律规范和法秩序。可见，实定法合理性假定是一种相对的合理性假定。按此假定，不可批判的是法律规范，而不是具体法条。

事实上，对于要不要通过不断修订刑法以扩大犯罪圈的问题，包括许多刑法教义学者在内的大多数学者都持肯定态度。如果所有实定法法条都绝对合理，为什么还要支持刑法修订以新增罪名，或改变某些罪名的入罪条件，或调整部分犯罪的法定刑、减少死刑？更何况，某些冠名以法教义学的研究，往往对规范体系中不符合其特殊预设的一些规范置之不顾，如“社会主义”“公有制”等类型的规范，往往被视为纯粹的意识形态表达而弃之如敝屣。〔7〕可见，法教义学阵营本身也非铁板一块，对实定法也是有选择地遵从，并且其选择标准及结果也是可疑可议的。再者，法教义学是以法律适用为中心而展开的。〔8〕实定法合理性假定的本意是法律适用不能以法外因素的影响、法律本身的不确定性或法律漏洞为由，另寻裁判理由。所以，

〔2〕 凌斌：《什么是法教义学：一个法哲学追问》，《中外法学》2015年第1期，第226页。

〔3〕 参见〔德〕古斯塔夫·拉德布鲁赫：《法教义学的逻辑》，白斌译，《清华法学》2016年第4期，第200页。

〔4〕 冯军：《刑法教义学的立场和方法》，《中外法学》2014年第1期，第173页。

〔5〕 许德风：《法教义学的应用》，《中外法学》2013年第5期，第938页。

〔6〕 〔德〕卡尔·拉伦茨：《法学方法论》，陈爱娥译，商务印书馆2003年版，第77页。

〔7〕 参见前引〔1〕，谢海定文，第93页。

〔8〕 参见陈兴良：《刑法教义学方法论》，《法学研究》2005年第2期，第41页。

批判法律和不依法裁判是两回事。

另一方面，实证研究的法律实践集体经验合理性假定，源于“法律是一种实践理性”的信念。按照这一信念，刑法不仅是犯罪行为作为犯罪定义对象的反映，也是法律制定者、适用者作为定义者自身各种主体性因素的彰显和表达。借助大范围、可重复性强的普遍法律实践，有助于最小化个体经验、主观潜见、偏好、局限甚至私利等主体性因素对法律适用的影响，使其消解在普遍法律实践的一般性中。也就是说，法律人的集体经验比个别经验更可能过滤掉个别利益局限和特殊的主体性偏好，更可能接近公认的普遍价值。这时，集体经验的合理性所彰显的是普遍价值本身的有效性。的确，受样本等因素影响，不同实证研究得出的数据不可能绝对相同，甚至相反。这个现象常被用来证明实证研究不可信。其实，如果把实证研究视为不断趋近集体经验的过程，那么，样本越大、分析工具越科学，实证研究的结论就越可靠。

不过，法律实践集体经验合理性假定也不是绝对的。比如，法律禁止任何人强奸女性，即使大多数（男性）法官对大多数所谓“婚内强奸”案的被告人都判决无罪，也不应据此将此类行为排除出强奸罪的犯罪定义。正如个别法条可能存在种种不足一样，法律实践集体经验的合理性也是相对的。所以，基于法律实践集体经验的研究，除了需要满足统计显著性、高相关度等形式要求以外，还需要进一步的法教义学分析，看其是否符合法律规范意义上的正当性要求。

既然两个合理性假定都不是绝对的，刑法教义学与实证研究之间就没有对立紧张的根本理由。当然，无论是实定法合理性假定，还是法律实践集体经验合理性假定，其相对性都是描述的相对性，与价值相对主义有着本质不同。价值相对主义否认超越多样性、地方性、时代性伦理价值的普遍价值的存在，认为不应该用普遍价值去规范人的行为。总之，承认实定法和集体经验的相对合理性，打破刑法教义学与实证研究之间的隔膜并非不可能。

## （二）效度与信度

法教义学与实证研究之间的另一层隔膜，是各自的技术短板：法教义学的短板在于信度有限，实证研究的短板在于效度不高。效度是指研究所选用的指标、维度的有效性。研究的结果与所要考察的内容吻合度越高，则效度越高。如果用卷尺测量人的体重，其效度就很低，尽管身高和体重的确呈正相关关系。信度是研究结果的稳定性及可靠性程度。如果测量人的身高所用的卷尺是用松紧带制成的，其信度就很低，尽管测量身高的确需要用卷尺。效度是前提，回答“研究什么”的问题；信度是保障，回答“怎么研究”的问题。

按此分法，有的刑法教义学理论具有较高的效度，却由于所使用的事实材料多为个案或小样本的直接观察，疏于可推论性、代表性交代而缺乏一定的信度。例如，常有人抱怨死刑适用标准不明确，于是有学者告诉我们，如果人们愿意把刑法关于死刑的规定作为一个整体来解释，就会发现刑法第121条、第239条和第48条之间的紧密联系，就会得出“只有故意致人重伤、死亡或者故意杀害他人的，或者以暴力、胁迫或者其他方法强取财物并且造成了与航空器遭受严重破坏相同程度的财产损失的，才属于极其严重的罪行”这一极为明确的结论。<sup>〔9〕</sup>这就是一个法教义学发现过程：从数个法条之间的逻辑关系建构法教义学内容，即死刑适用标准的明确性。不过，仅仅三个法条的描述，显然不能直接代表中国刑法中几十个死刑罪名之间的全部逻辑关系。

〔9〕 参见冯军：《刑法学研究之检讨与反思》，《法学研究》2013年第1期，第19页。

另一方面,不少刑法实证研究装备精良、数据分析工具先进,却不知到底在研究什么问题,缺乏应有的效度。这些研究借助大量图表数据的堆砌来描述刑法现象,但对某某率或某某系数的意义缺乏交代,让人无从把握这些数据描述与刑法教义学概念、理论有何关系。有的研究沉迷于复杂的数学模型和量化描述,以为选用的数学工具越复杂、越深奥,其结论就越科学。其实,研究的工具理性和程序理性是为理论研究和认知推进服务的,没有实体理论的准确把握,再复杂的计算方法也没有法学意义。

解决上述短板的关键在于法教义学与实证方法合作,对抽象概念、理论命题进行操作化处理。只有当概念、理论命题转换为可度量的变量、指标以及可检验的工作假设时,才可能实现刑法教义学与实证研究之间的无缝对接。否则,便可能形成刑法教义学与实证研究之间谁看谁都别扭的局面。刑法中,大到公正、正义、均衡、法益、秩序、该当、违法、有责,小到行凶、防卫限度、入户、性行为、损失,都需要转换处理。例如,按照刑法规定,金融机构工作人员造成一定程度的损失便可能构成相关犯罪。而在金融业,“损失”只有先拆分为规模、确定性、持续性、致命性等几个维度,并进一步分别下降到损失金额大小、不良资产的五级分类法定位、逾期时间长短、可支配资金额度的占比等几个具体指标后,才可能弄清一组样本的损失与相关刑事责任之间的关系以及相关刑法规范的适用实效。再如,“刑法的明确性”是个典型的法教义学问题,黑静洁就将“明确性”这一传统刑法学中最不明确的概念,<sup>[10]</sup>操作化为定量因素、兜底条款、空白刑法规范和过失犯等指标,并据此对中外刑法进行了比较。结果发现,中国刑法的法定犯和经济财产类犯罪相关规定的明确性程度较低。据此,他建议完善附属刑法以提高刑法的明确性程度。<sup>[11]</sup>可见,用实证方法研究刑法教义学问题并非不可能。

应当承认,即使以上两层隔膜可以消除,实现刑法教义学与实证研究的深度合作还有以下困难:刑法教义学者的经验研究技能训练不足,实证研究者的法教义学知识匮乏;法律规范文本和司法裁判文书中对同类刑法现象的表述不一,裁判的说理程度不高;反映案件真实信息的卷宗基本不公开,某些数据信息仍有保密要求;从法律文本自然语言中提取标准化数据信息的技术能力有限,缺乏高质量的数据信息共享平台,等等。但是,仍应相信,克服这些障碍将有助于提高中国刑法研究的水平。

## 二、刑法研究需要法教义学与实证方法的新型合作

可能合作不等于必要合作。本来,法教义学和实证研究已经从各自角度为中国刑法研究做出了贡献,为什么还要提倡二者的合作?理由主要如下:

### (一) 法律是一种实践理性

在康德看来,理性分为理论理性和实践理性。理论理性基于人与自然的关系,是人的认知思维能力的运用。实践理性基于人与人的关系,是人的道德思维能力的运用。在康德哲学中,这两种理性的区别在于:第一,对象不同。理论理性的对象是自然,实践理性的对象是人的自由。第二,起点和终点不同。理论理性从直观、感性开始,达至理论;实践理性从理论开始,

[10] 参见刘艳红:《开放的犯罪构成要件理论研究》,中国政法大学出版社2002年版,第306页。

[11] 参见黑静洁:《刑法明确性原则的恪守程度——刑法明确性原则的衡量指标及其实证考察》,《政治与法律》2016年第11期,第93页。

走向感性、经验。康德认为，理论理性从感性开始，进展到概念、原理；相反，实践理性并不和客观对象打交道，而是和意志打交道。<sup>[12]</sup> 第三，具有不同的规定性。理论理性主要回答“是什么”的问题；实践理性服从应然的绝对命令，主要回答“应当是什么”或“应当如何”的问题。第四，两种理性的内容不同。理论理性是主体的先天认识能力，实践理性是主体的先天欲望能力。康德指出，对道德行为而言，不是对象而是意志的法则，才是行动的规定根据，所以，实践理性的唯一客体就是那些善和恶的客体。<sup>[13]</sup> 可以说，理论理性是真理论的主要论题，而实践理性是价值论的主要论题。

对照以上几点，面对人的自由、权利、欲望、应当、善恶，法律显然属于实践理性的范畴。专注普遍法律实践的实证研究自然也离不开实践理性。问题是，该如何理解法教义学与实践理性的关系。据赫尔贝格的考证，“教义”一词其实最早出现于医学文献。在《希波克拉底全书》中提到了医学“实践学派”与“理论学派”之争，“实践学派”否认医学能得出任何一般命题，认为有的只是一个个特定病例，而“理论学派”认为可以从最一般的原则出发演绎推导出一般命题（假定）。但随后，有一部分“实践学者”开始寻求具体病例之间的共性。同时，许多“理论学者”也开始意识到，有必要通过归纳来对病例的具体观察作一般化总结。到了大约公元前400年左右时，出现了走第三条道路的学者，被称为“教义学者”，并将这种类型的一般原理或命题命名为“教义”。<sup>[14]</sup> 可见，“教义”原本是经验归纳与理性抽象相结合的产物。而且，法教义学与实证研究相和，在我国也已有尝试。<sup>[15]</sup> 所以，刑法教义学应回归到其原本意义上来使用“教义”一词。这样，刑法教义学与实证研究至少在语源学的意义上具有亲缘关系。

## （二）法律规范隐身于大量实定法文本中

法律规范不是法律条文的任意堆放，而是法律内在结构形成的逻辑关系。发现这种关系，将隐身于法律条文中的刑法规范（教义）提取出来，既是法教义学也是实证研究的分内之事。其中，演绎推理和归纳方法都是发现法律规范的必要方法。例如，张明楷关于法益保护的论证，就以刑法为对象，通过观察、归纳大量法律条文，说明这些法条背后的共性是法益保护对刑法的指导意义。<sup>[16]</sup> 的确，演绎法擅长发现法律中未言明的规范。例如，刑法只规定盗窃并使用信用卡的，定盗窃罪，而未规定诈骗、抢劫、抢夺信用卡并使用的应如何定罪。实际上，既然盗窃并使用信用卡的定盗窃罪，而诈骗、抢劫、抢夺等行为与盗窃具有相当性，那么也可以认为诈骗、抢劫、抢夺信用卡并使用的，应以信用卡取得行为分别定诈骗罪、抢劫罪、抢夺罪。此即作为无形之法的法教义积极主动找法的过程，<sup>[17]</sup> 也是法教义学演绎逻辑应用的典型范例。

与此不同，归纳法擅长发现明文规定的法条之间的关系或规律。例如，1979年刑法中有27个死刑罪名，死刑罪名最多的是反革命犯罪，55.6%的死刑分布在“反革命罪”一章中。

[12] 参见〔德〕康德：《实践理性批判》，邓晓芒译，人民出版社2003年版，第122页。

[13] 同上书，第79页。

[14] 参见雷磊：《法教义学观念的源流》，《法学评论》2019年第2期，第42页。

[15] 我国台湾法官就曾以犯罪统计数据为根据，证明禁止某些有犯罪前科者办理小客车营业执照可能“违宪”。这里，一种宪法教义学学说已经容纳了特定的经验知识。参见雷磊：《法教义学的基本立场》，《中外法学》2015年第1期，第203页。

[16] 参见张明楷：《新刑法与法益侵害说》，《法学研究》2000年第1期，第19页。

[17] 参见陈兴良：《立法论的思考与司法论的思考——刑法方法论之一》，《人民检察》2009年第21期，第8页。

1997年刑法中共有68个死刑罪名,死刑罪名最多的变为破坏社会主义市场经济秩序犯罪,23.5%的死刑分布在经济犯罪中。到了2011年,刑法分则中死刑罪名减少到55个,死刑罪名最多的是危害公共安全犯罪,有25.5%的死刑分布于此。死刑分布的这些变化,代表了刑事政策重点的两次重大转移。1979年,刚刚结束“文革”的中国,许多人仍然相信“阶级斗争”的存在。经过近二十年的改革开放,大规模经济建设和随之而来的各类新型经济关系,从各方面影响着社会生活。进入21世纪的最初十年,国际、国内恐怖犯罪和各类危害公共安全的犯罪,越来越成为社会控制的焦点。反革命犯罪、破坏市场经济秩序犯罪和危害公共安全犯罪,先后分别代表了几十年来最为危险的几类犯罪。因此,死刑的分布变化既反映了社会危险的前后变化,又不断指导乃至拉着社会惩戒资源的主动配置。这种从大量法条中寻找规律的归纳研究,也是对本国法律文本的深耕细作,是另一种形式的向实定法致敬。总之,透视法条背后的关系、规律,需要法教义学和实证研究的互补与协作。

### (三) 法教义学与实证研究交集于法的有效性

法的有效性的应然面是法的效力,其实然面是法的实效。效力是实效的前提,实效是效力的实现,法治是效力不断转化为实效的过程。<sup>[18]</sup>法教义学研究法的应然有效性,为法律实证研究提供目标与方向。法律实证研究专注法的实然有效性,研究法律有效性的实际程度、范围,是法教义学的必要补充。例如,劳佳琦对一千多个累犯样本进行的实证研究发现,目前我国普通累犯制度的隔离机制低效,威慑机制无效,在减少犯罪、防卫社会方面的综合实际效果远远低于预期;累犯从严有效论可能只是个神话。<sup>[19]</sup>在美国,有学者研究“不退让法”如何影响枪支致死致伤的数量。他们用倍差法分析了州一月级别的数据,发现“不退让法”导致了枪支杀人事件的增加。“不退让法”每月至少导致30人被杀,以及更多的枪伤住院事件,因而质疑“不退让法”是否对维护公共安全起了正面作用。<sup>[20]</sup>

法的有效性的核心是法律规范对相关主体的约束力,包括对法律适用者的约束力和对普通公民的约束力。其中,对法律适用者的约束力,是法教义学与实证研究之间交集的交集。因为犯罪控制与其说是针对犯罪的控制,不如说是对犯罪控制的控制,强调对犯罪控制者的行为有所约束。否则,不仅会有更多的行为被定义为犯罪,而且可能引起某些犯罪。所以,刑法适用的真正意义并不只是有利于被告人,更在于最大限度地限制法律适用者刑罚权的行使过程。

在这方面,法教义学和实证研究的作用都不可替代。一方面,法律规范的实效与规范自身有关。例如,在刑法修正案(六)之前,刑法第187条规定了用帐外客户资金非法拆借、发放贷款罪,但犯罪统计显示该规定的实施效果并不理想。原因是,按照原刑法规定,如果以其他个人目的,或者将不入帐的资金用于其他用途,或者未造成重大损失,均不符合“以牟利为目的”,或者“将资金用于非法拆借、发放贷款”,或者“造成重大损失”等条件,因而都不构成本罪。所以,少有定罪案例,该规定对银行资金安全的保护效果并不理想。对这类问题,就需要法教义学积极参与下的刑法修订来加以解决。于是,才有了后来刑法修正案(六)规定的吸收客户资金不入帐罪,改变了相关犯罪的入罪条件。另一方面,法律规范的实效还与司法地方性、经济、文化、政策等法律规范外因素有关。王剑波关于量刑地方性差异的实证研

[18] 参见刘焯:《法的效力与法的实效新探》,《法商研究》1998年第1期,第79页。

[19] 参见劳佳琦:《我国累犯从严实效之实证研究》,《中外法学》2014年第6期,第1648页。

[20] See McClellan, Chandler, and Erdal Tekin, *Stand Your Ground Laws, Homicides, and Injuries*, 52 (3) *Journal of Human Resources* 621-653 (2017).

究发现，刑法修正案（九）和“两高”司法解释确立的“数额+情节”的弹性定罪量刑标准，难以消减受贿罪量刑的地区差异。<sup>[21]</sup>可见，法的有效性研究既离不开法教义学，也不能没有实证研究。

综上，法教义学和法律实证研究皆为实践理性范畴的一部分，二者具有良好的合作基础。法律规范隐身于大量实定法文本中，两种方法论的合作有助于法律规范地发现。法的有效性集法律规范效力和实效性于一身，没有法教义学的实证研究或者没有实证研究的法教义学，都无助于法的有效性的实现和检验。所以，法教义学实证研究很可能代表了中国刑法研究一个值得探索的方向。

### 三、刑法教义学实证研究

刑法教义学实证研究是以犯罪论教义学、刑罚教义学和刑法各论（罪名）教义学理论和实践为对象的实证研究。限于篇幅，现以刑罚教义学实证研究为例，解释何为刑法教义学实证研究及其基本特征。

#### （一）刑罚教义学实证研究

刑罚教义学是以法定刑配置、刑罚适用的基本原则、规范及实践为内容的教义学化刑法知识。死刑适用原则的法教义学化知识，是刑罚教义学的重要组成部分。所以，死刑适用原则的实证检验，属于刑罚教义学实证研究。

根据刑法第48条的规定，死刑只适用于罪行极其严重的犯罪分子；对于应当判处死刑的犯罪分子，如果不是必须立即执行的，可以判处死刑同时宣告缓期二年执行。这个规定中有两条界限：一是如果不满足“罪行极其严重”的要求，就不能进入死刑圈，只能适用自由刑或其他刑罚；二是如果罪行极其严重但“不是必须立即执行”的，就只能留在死刑缓期执行圈。同时，刑法中有27个故意犯罪的罪状中含有致人死亡的结果。除去虐待罪等，常见的暴力致人死亡犯罪共有八个罪名：故意杀人罪、故意伤害罪、强奸罪、绑架罪、抢劫罪、放火罪、爆炸罪、投放危险物质罪。由此形成的案件可称为“命案”，刑法对这些生命犯罪都规定有死刑及严格的死刑适用条件。这八个罪名的死刑适用条件和刑法第48条的内容连起来看，其内在逻辑就是必须严格限制死刑的适用。因此，“严格限制死刑适用”可以视为一项暗含在上述九个法条中的刑罚教义学原则。

这一刑罚教义学原则的规范意义还具有一种无形的延展性：最极端的恶就是非法暴力剥夺他人生命的犯罪，最极端的社会否定性评价就是依法剥夺犯罪人生命的死刑，这两个剥夺是其他一切剥夺的边界。草率对待他人的生命，就难免草率对待他人的其他权利——暴力剥夺他人生命的个人和国家概莫能外。国家如何对待非法剥夺他人生命的人的生命，将传导至刑事司法乃至社会的各个角落，因而是了解中国刑法最具代表性的切入点。严格限制死刑适用，对被告人、受刑人、潜在的犯罪人乃至全社会，将是一种最好的宣示：暴力应该是最需要严格限制的社会反应手段。否则，惩恶未必扬善，甚至可能默许暴力。从这个意义上说，命案之罪的死刑适用是中国刑法的“第一张骨牌”。

如何安放这张骨牌，集中了许多种二元化理论：有的强调客观罪行，有的强调主观罪恶；

[21] 参见王剑波：《我国受贿罪量刑地区差异问题实证研究》，《中国法学》2016年第4期，第245页。

有的看重罪行严重性的影响,有的看重犯罪人人身危险性的作用;另外还有结果无价值论与行为无价值论、报应论与预防论,等等。如果同时从“罪大”和“恶极”两个视角看,便有四种组合:罪大且恶极、罪不大且恶不极、罪大但恶不极、罪不大但恶极。前两种组合比较极端,其结果更可能是死刑立即执行或者排除在死刑圈之外。现实中后两种组合才是大概率事件,法律适用会面临许多复杂情形。结果,到底是杀人偿命还是免死,从(犯罪)现场到(刑罚执行)刑场之间,到底发生了什么,严格限制死刑适用的实效到底如何,就需要进行刑罚教义学实证研究。

研究首先将“严格限制死刑适用”的概念操作化为三组变量:第一,有多少死罪适用了死刑。第二,“罪行极其严重”的判断到底受哪些因素的影响。第三,“不是必须立即执行”的判断到底受哪些因素的影响。其检验逻辑是:其一,死刑适用率越低,则越大程度上证否“杀人偿命”的假定,从而接近事实上废除死刑的状态,<sup>[22]</sup>也才越有可能批驳所谓“中国滥用死刑”的猜测。其二,影响死刑适用的实际因素如果更多是客观实害因素,则支持报应论或者结果无价值论;如果更多是主观因素、人身危险性因素,则支持预防论或者行为无价值论。其三,限制死刑适用的因素、情节对死刑适用的影响越稳定、越具有规律性,说明越符合同恶同罚的要求。这三组变量及其检验逻辑的确定,是实现抽象的“严格限制死刑适用”与具体经验数据之间转换的技术关键。

接下来,我在北京市高级人民法院和最高人民法院对部分长期从事死刑适用工作的资深刑事法官进行了访谈。之后,又在中国裁判文书网350万刑事案件中提取全部八种命案且有致死结果的案件近8万个(79904个),作为研究样本。<sup>[23]</sup>由于命案之罪未必都有致死结果,仅选取有致死结果的命案作为样本,才可以更清楚地看出,什么样的“杀人”才会偿命。

研究先将样本中未适用有期徒刑的近3万个判处无期徒刑及以上刑罚的案例分为两组:第一组为判处死缓及无期徒刑的样本,共26332个,其中死缓10777个、无期徒刑15555个。对该组内两类案例进行比较,目的是发现何为“罪行极其严重”的实际判断标准。第二组为判处死刑立即执行及死缓的样本,共14035个,其中死刑立即执行3258个、死缓10777个。对该组内两类案例进行比较,目的是发现何为“不是必须立即执行”的实际判断标准。然后,对两组样本分别进行二元Logistic回归分析。由于回归分析的原理是在控制其他变量影响的情况下观察某个因素的作用,所以,如果发现某个量刑情节对量刑结果的影响满足统计显著性要求,就意味着该情节对死刑适用的某个界限区分具有普遍意义。运行量化分析过程后有如下发现结果:

结果一:本研究范围内命案致死样本的死刑立即执行适用率为4.1%,死缓适用率为13.5%,无期徒刑适用率为19.5%,有期徒刑适用率为62.7%。另有命案致死却无罪及免于

[22] 根据联合国秘书长提交联合国有关机构的报告,世界各国和地区分为完全废除死刑的国家和地区、废除普通犯罪死刑的国家和地区,以及事实上废除死刑的国家和地区。第三类国家和地区在法律中保留死刑并可能作出死刑判决,但在统计时间之前的至少10年里未实际执行过死刑。在秘书长报告中,从2005年开始,在统计时间之前的10年里执行过死刑,但已作出国际承诺、正式宣布暂停使用死刑的,也被列为事实上废除死刑的国家和地区。参见孙世彦:《从联合国报告和决议看废除死刑的国际现状和趋势》,《环球法律评论》2015年第5期,第119页。

[23] 关于样本的可推论性,需要作出特别说明。中国裁判文书网公布的死刑案件并非死刑案件的全部。因此,我们无法精确报告上述样本结构的代表性。但是,首先,本研究的目的是观察死刑适用过程和标准,只要抽样过程科学、客观,就符合研究目的的要求。其次,虽然不是全样本,但近8万个样本至少是大样本。再次,即使上传的数据是有选择的,其选择标准也不大可能针对本研究的理论关注来确定,所以可以排除样本系统性偏差的可能性。



刑事处罚样本 26 例和 202 例。即 80% 以上构成犯罪且致命的命案，最终的刑罚是监禁刑。再考虑到实践中绝大部分被判处死缓的犯罪人最终可能减为无期徒刑，监禁刑的适用率应该更高。这意味着，“杀人偿命”在中国已成为极小概率的反应方式。<sup>[24]</sup>对致人死命犯罪的反应尚且如此，其他罪刑关系的处理又该当如何？

结果二：从罪名看，只要是故意伤害罪（致死），不论是对“罪行极其严重”的判断来说，还是对“不是必须立即执行”的判断来说，选择相对较轻的量刑结果都是大概率事件。证据是，在第一组样本中，“是否故意伤害致死”的标准化回归系数为  $-0.887$ ，<sup>[25]</sup>在所有影响因素中排序第一；在第二组样本中，“是否故意伤害致死”的标准化回归系数为  $-0.640$ ，也在所有影响因素中排序第一。

结果三：对是否罪行极其严重的判断来说，致死数量的影响小于重伤数量；而对是否死刑立即执行的判断来说，致死数量的影响大于包括重伤数量在内的许多变量。证据是，在第一组样本中，致死数量和重伤数量的标准化回归系数分别为  $0.033$  和  $0.125$ ；而在第二组样本中，致死数量和重伤数量的标准化回归系数分别为  $0.608$  和  $0.171$ 。尤其有趣的是，研究将致死数量分为“是否致三人及以上死亡”作为控制变量，重新运行 Logistic 回归分析过程后发现，在致死三人及以上的样本范围内，死刑立即执行预测的准确率高达  $94.3\%$ ，其余样本死刑立即执行预测的准确率也高达  $86.1\%$ 。而且，H-L 值<sup>[26]</sup>为  $0.811$ ，说明模型预测效果较好。控制后的回归分析结果是，大部分原有自变量的显著影响都消失了，这说明在最终决定是否死刑立即执行时，致三人及以上死亡基本上是导致死刑立即执行的最重要影响因素。这个事实可能是支持客观主义、结果无价值论最有力的证据。但还是不能据此认为，主观因素对死刑适用没有影响。我们不能想象，一个故意杀死三人及以上的犯罪人主观上恶性不大，没什么人身危险性，判处死刑，只是因为客观上致死人数太多。

结果四：来自犯罪人的因素，也对死刑适用构成显著影响。前科劣迹及累犯在是否进入死刑圈和是否死刑立即执行的判断中，都呈显著趋重影响。犯罪后采取补救措施仅对是否罪行极其严重构成显著负相关关系，而自首、立功、精神病人等几个因素，对两个界限的判断都呈显著负相关关系。其中，尤以自首对是否死刑立即执行的判断影响力度最大，标准化回归系数高达  $-0.562$ ，其影响力位居第三，仅次于故意伤害致死和致死数量两个变量。可见，在成立自首的情况下，被判处死刑立即执行的比较罕见。这些因素大都发生在犯罪前或犯罪后，若用犯罪本身的行为无价值或结果无价值来概括，似乎都过于简化，忽略了大量显著影响法官裁判活动的重要信息。

结果五：在死缓与无期徒刑的抉择中，只要取得被害方谅解，就很可能判处无期徒刑而不

[24] 对此，最可能的一个质疑是，这个比例是否忽略了死刑案件上网受到某种限制的影响。的确，我们没有根据说，死刑案件上网规模与命案上网范围受到同比例控制。不过，毕竟尚无根据说死刑案件未上网的实际规模到底有多大，本研究至少为该问题的研究提供了一个相对较大的观察样本。

[25] 标准化回归系数显示自变量（量刑情节）对因变量（量刑结果）的影响力度大小和方向，越接近 1，说明该自变量对解释因变量的贡献率越大，反之越小；其正负值表示作用的方向。统计软件 SPSS 的 Logistic 程序不提供标准化回归系数，其 B 值只是非标准化回归系数。因此，本研究先将定类、定序变量转换成虚拟变量，然后对所有变量进行标准化处理，运行 Logistic 程序后，其 B 值便是标准化回归系数。下同。

[26] 霍斯默-莱梅肖拟合优度检验（Hosmer-Lemesho, H-L 检验）用于评价模型是否充分利用了现有信息，最大化地拟合模型，解释了模型的变异。若  $P > 0.05$ ，则可说明模型拟合优度效果较好；若  $P < 0.05$ ，则说明模型构建效果欠佳。

是死缓，标准化回归系数为 $-0.267$ 。对死缓还是死刑立即执行的抉择来说，取得被害方谅解的，更可能被判处死缓而非死刑立即执行，标准化回归系数为 $-0.534$ 。而是否取得被害方谅解，至少与两个因素有关：一是被害过错因素。在刑法学中，通常认为被害过错的存在降低了加害人行为的主观恶性和可责性。因为被害人的过错既可能导致被告人的动机值得宽恕，也可能导致被告人的期待可能性减少，故成为减少责任刑的情节。<sup>[27]</sup> 在本研究的近8万个命案样本中，有被害过错的样本共12626个，占比15.8%。只要有被害过错，在死缓还是无期徒刑的抉择中，就更可能判处无期徒刑而不是死缓；在死刑立即执行还是死缓的抉择中，就更可能判处死缓而非死刑立即执行；两者的标准化回归系数均为 $-0.217$ 。实践中，取得被害方谅解往往与被害人过错高度相关，后者更可能成为被害方谅解的心理基础。二是赔偿被害方损失。在适用死刑且达成被害方谅解的样本中，赔偿被害人的占比为58.3%。相比之下，赔偿被害人却未达成谅解的，在赔偿样本中占比高达93.5%。这说明，大部分赔偿不会达成谅解，而谅解也不全是赔偿所致。但毕竟，拒绝赔偿的与死刑立即执行之间呈显著正相关关系，尽管其解释力有限，标准化回归系数仅为0.091。上述谅解、赔偿、被害过错等因素或多或少与罪行以外或犯罪人以外的因素有关，也无法简单地从犯罪本身是客观还是主观的二元区分来概括。

结果六：是否罪行极其严重以及是否必须死刑立即执行的判断，还与案件审理过程中的程序性因素甚至案外因素有关。研究发现，被羁押时间长短与死刑立即执行之间呈高度负相关关系，标准化回归系数为 $-0.399$ 。这也许是因为，羁押时间越长，通常说明案件疑难程度越大，最终是否判处死刑立即执行就越需要慎重决断。如果证人出庭，也明显降低了死刑立即执行的概率，标准化回归系数为 $-0.173$ 。尽管理论上证人出庭作证既可能由辩方申请也可能应控方要求，但实践中，绝大多数情况下是辩方向合议庭申请证人出庭作证。因此，证人出庭很可能对被告人有利。研究还发现，在是否死刑立即执行的判断中，如果被告人是少数民族，被判处死刑立即执行的概率明显降低，标准化回归系数为 $-0.105$ ，这反映出法官对判决的社会效果的某种预估。这些分布特征意味着，法官并不是仅仅用罪行、犯罪人等要素来诠释刑法第48条的规定。

该研究中，从“杀人偿命”到“接近事实上废除死刑”之间，从某因素对死刑适用毫无影响到回归系数为1之间，有许多可能的相对位置。正是这些相对位置的客观描述，使“严格限制死刑适用”这一刑罚教义学原则的实效，在不同时间、空间实现了可比性，基本上证否了“杀人偿命”的说法。而且，刑法理论中的某些二元化思维，只是刑法现象的不同侧面在人们头脑中的反映，也是人们在观念上对刑法世界的主观建构。没有这些主观图式，人们就很难将枯燥的刑法规则与日常经验或其他社会科学理论联系起来。上述刑罚教义学实证研究证明，没有哪种二元化理论可以单独概括普遍的刑罚适用实践。量刑情节影响的稳定性，也因变量的不同而不同，无法一概而论。

## （二）刑法教义学实证研究是一种描述性批判

如果把法教义学视为内容，把实证研究视为形式，那么，刑法教义学实证研究就是一种以刑法教义学命题为理论假设，以刑法文本信息为事实材料的假设检验过程。一般认为，只有当检验结果全面或部分证否某个理论假设时，才认为发现了未知。如果一个假设本身就不可证否，或者未能证否某个可证否的假设，就都算不上标准的实证研究。所谓描述性批判，就是以

[27] 参见张明楷：《责任刑与预防刑》，北京大学出版社2015年版，第286页。

事实数据描述证否某个刑法理论假设。在前述“第一张骨牌”研究中，通过死刑立即执行适用率、死缓适用率、无期徒刑适用率、有期徒刑适用率等几个指标的数据检验，基本上证否了“杀人偿命”的说法。这意味着，过分相信死刑作用的理论是不能成立的，也不存在滥用死刑的趋势。其实，这种描述性批判对中国刑法研究的影响越来越明显。例如，形式理性与实质理性之争，是刑法学中的重大理论问题之一。王政勋对大量法官群体和在校学生群体的问卷调查结果进行比较，结果发现，法官在理解、解释和适用法律时并没有采取主观解释、形式解释的立场，而是坚持实质理性的追求。<sup>[28]</sup>赵书鸿的实证研究发现，尽管理论上会认为量刑既要考虑报应又要考虑预防及其他各种因素的影响，但刑罚裁量其实是一个以确定行为严重性为导向，仅对有限变量进行笼统评估的过程。<sup>[29]</sup>

这种证否式的描述性批判，会丰富我们关于实定法的理解。例如，从是否侵犯生命的犯罪与是否配有死刑两个角度进行划分，刑法分则中的罪刑关系便有四种排列组合：有死刑的生命犯罪、无死刑的生命犯罪、有死刑的非生命犯罪、无死刑的非生命犯罪。所有尚未废除死刑的刑法都可以放入这个框架中进行分析。结果，有的刑法中有死刑的非生命犯罪可能很少，甚至根本不存在，而有的则可能很多。只比较一下刑法修正案（五）时的中国刑法与今天的中国刑法两个文本，就可以从统计结果中看出实定法本身的这种结构性变化：有时，大部分死刑分配给了非生命犯罪，尽管这种情况已经有所改变，但毕竟尚未消失。如果说出这个事实就是在批判刑法的话，不仅刑法的实证研究只好认账，就连刑法教义学者也不会反对说出这个事实。否则，就无法解释为什么连续两次减少死刑罪名恰好全部针对那些有死刑的非生命犯罪。这就是让数据自己出来说话的描述性批判，即通过说出事实——往往是基于大样本观察描述的事实而非“拿鞋找脚”式挑选来的个别事实——揭示未知的关系、规律。所以，这个意义上的批判，是一种优雅的批判、一种没有主观预设的批判。

需要说明的是，对二元性的竞争性理论来说，证否与证成是相对的。例如，围绕死刑的存废，历来存在死刑有无威慑力两种对立的学说。吴雨豪采用“准实验设计”方法所进行的死刑威慑力实证研究就证实，2007年最高人民法院收回死刑复核权导致中国的死刑实际执行数量明显减少。其结果，死刑政策的限缩并未导致犯罪率上升，从而证否了死刑具有威慑力的观点。<sup>[30]</sup>这里，死刑保留论的证否其实意味着死刑废除论的证成，反之亦然。可以看出，由可靠的事实自己说话，是一种不动声色的冷静的批判。

### （三）刑法教义学实证研究的描述对象是刑法现象大样本

天下没有两片一样的树叶，负责任的研究者不会轻率地用个别实例直接推论总体。所以，刑法教义学实证研究所描述的，不是单一法条或个案，而是刑法现象大样本，包括一定数量规模的司法裁判，或者多个法律条文以及调查获得的数据。用大样本描述进行学术批判，可以更好地防止小样本自身或个例选择者的特殊性及其主观偏好的影响，因而其学术批判的信度更高。前述“第一张骨牌”研究的样本，是中国裁判文书网350万刑事案例中所提取的全部八种命案且有致死结果的近8万个个案。如果没有可靠性与其大体相当的样本，推翻相关结论就可能需要另辟途径。这种严肃的大样本研究还有许多，如梁根林等人的中国死刑民意调查样本多达

[28] 参见王政勋：《从图式理论看刑法解释立场：一个实证的考察》，《中外法学》2009年第3期，第358页。

[29] 参见赵书鸿：《论刑罚裁量的简洁化——量刑活动的经验性研究结论》，《中外法学》2014年第6期，第1626页。

[30] 参见吴雨豪：《死刑威慑力实证研究——基于死刑复核权收回前后犯罪率的分析》，《法商研究》2018年第4期，第93页。

3万多个,是目前该类研究最大的样本;〔31〕劳佳琦的累犯研究样本为10万个;〔32〕赵兴洪的缓刑研究样本为130万个;〔33〕在中国裁判文书网全面上线前,蔡曦蕾的量刑失衡研究样本也达到3.3万多个。〔34〕这些研究的相关结论至今尚未遭遇有力的挑战。

大样本描述必然带来学科交叉与融合。与学科内的论战相比,学科交叉融合所形成的挑战同样很有意义。法教义学实证研究本身就是法教义学、统计学、计算机信息技术三者的交叉融合。其中,法教义学是内容本体,统计学是描述方法,计算机信息技术是事实信息挖掘、处理工具。比如,基于信息科学的语义分析工具,可以对刑法文本进行还原分析,将出现频率较高的八百多个动词归纳为三类:“放火”“破坏”“持械”“残害”等归结为暴力;“盗窃”“贪污”“盗掘”等归结为偷窃;“骗取”“煽动”“谎报”“骗购”“引诱”等归结为欺骗。这就为罪名教义学实证研究创造了类型学量化分析的条件。可见,对于“摆事实,讲道理”中的“摆事实”,法教义学实证研究用具体的数量规模、比例关系、相关系数、回归系数等取代“举例来说”“多数情况下”“通说认为”“往往”等模糊性诱导或强制性暗示用语。这种“摆事实”基础上的“讲道理”更具说服力,其结论也更具一般性和可推论性。

一个可能的质疑是,司法裁判和社会调查结果固然是刑法现象大样本,而实定法文本可否作为描述对象,对其进行描述性批判?回答是肯定的。劳东燕曾对中国刑法、刑事诉讼法、刑事司法解释、刑法规范性文件进行筛查检索,发现有44个“推定”,并以此为样本进行实证分析。结果发现,刑事推定在放松控方证明要求的同时又将存疑风险转移到被告人身上,背离了排除合理怀疑的证明标准,直接危及无罪推定原则所保护的权益。〔35〕可见,刑法及司法解释条文大样本同样具有可证否性——法律本身不应是“法外之地”。

#### (四) 刑法教义学实证研究的结论是一种概括性批判

法教义学实证研究一般不直接否定个别法条、规定或生效裁判的正当性,其研究结论往往是一种概括性证否、批判。因为不论是法教义学的实定法合理性假定,还是实证研究的法律实践集体经验合理性假定,都非常看重法的一般性和普遍约束力。在“第一张骨牌”研究中,有的统计分析结果意味着对报应论或责任刑理论的证否,而有的是对预防论或预防刑理论的证否;有时你会从实证研究结果中看到结果无价值论的局限,有时则会看到行为无价值论的不足。最终人们会意识到,不论二元刑法理论中的哪一种理论,都无法单独解释大范围刑法现象。与其纠缠于这些二元化思维的桎梏,不如去发现某些不约而同的普遍做法。德国学者阿尔布莱希特的实证研究发现,在德国和奥地利,刑罚裁量的普遍水平位于法定刑幅度的三分之一以下。〔36〕这与中国常见犯罪量刑的实证研究结果相互印证,从而证否了量刑基准的法定刑幅度中间线说。〔37〕这种结论本身可能不具强制力,却为规范的制定、修订、司法实践或法教义学理论的完善,提供了可靠的事实依据。

刑法学上通常把不作为犯罪分为纯正和不纯正两种,问题是,这种划分周延吗?为了回答

〔31〕 参见梁根林、陈尔彦:《中国死刑民意:测量、解构与沟通》,《中外法学》2020年第5期,第1256页。

〔32〕 参见劳佳琦:《累犯从严量刑适用实证研究》,《中国法学》2016年第1期,第247页。

〔33〕 参见赵兴洪:《缓刑适用的中国图景——基于裁判文书大数据的实证研究》,《当代法学》2017年第2期,第46页。

〔34〕 参见蔡曦蕾:《量刑失衡的克服:模式与选择》,《中外法学》2014年第6期,第1596页。

〔35〕 参见劳东燕:《认真对待刑事推定》,《法学研究》2007年第2期,第21页。

〔36〕 参见〔德〕汉斯·约格·阿尔布莱希特:《重罪量刑》,熊琦等译,法律出版社2017年版,第322页以下,第523页。

〔37〕 参见白建军:《裸刑均值的意义》,《法学研究》2010年第6期,第135页。

这个问题，先穷尽刑法分则中所有表述为“拒绝……”“拒不……”“不履行……”的罪名共41个，圈定法定不作为犯的全样本，并严格依其法条表述分为充要不作为犯、必要不作为犯、选择不作为犯、混合不作为犯四种。与法定不作为犯越相近似的犯罪，才有越大的机会落入不作为犯的范围。按此标准对刑法分则条文进行排查，结果发现：所谓不纯正不作为犯的理论概念不具备划定刑法义务边界的功能，基本上没有提出哪些犯罪不可能由不作为构成的依据，这给刑法义务的任意扩张留下了空间。<sup>[38]</sup>这一研究还说明，刑法教义学实证研究完全可以实现演绎逻辑与归纳逻辑的并用。借助这种方法，研究从法律条文的集合中证实了“法定不作为犯的成立都以行为人对危害结果有所预见而不仅是应当预见为前提”等多个满足统计规律的逻辑关系。若有行为仅仅因行为人“应当预见”而被归入不作为犯罪，就意味着突破了实定法中相应逻辑关系的约束。可见，概括性批判既发挥了实证研究的长处，又满足了规范分析的一般性要求。

#### （五）刑法教义学实证研究结果的相对性

不同于个案分析，描述性批判不一定是彻底否定，也不是只有证实或证否两种结果。法教义学实证研究的结果具有明显的相对性，位于完全证实与彻底证否之间的许多相对位置上。按照“第一张骨牌”的研究结果，把死刑适用率和监禁刑适用率放在一起对照，限制死刑适用的实际程度就一目了然。用回归系数表示各种量刑情节的实际影响程度，有的接近0，有的接近1；有的影响力相对较小，有的则较大；在多种量刑情节的共同作用中，有时责任刑情节的影响力较大，有时预防刑情节起主要作用；把所有已知情节的共同影响加在一起，也无法百分之百地解释量刑结果，总会或多或少存在一定的未知空间。这种相对性描述比非黑即白式的批判更符合实际。在这类研究看来，任何理论都有自己的优势和局限，重要的不是谁的理论绝对正确，而是何种理论对哪类问题相对更具解释力。比如，关于谅解与赔偿对量刑的影响，持否定立场的理论可能认为，赔偿、谅解的引入实际上是拿钱买命赎刑，违背了法律面前人人平等原则。而高通的研究发现，赔偿对主刑量刑结果和缓刑适用均有重要影响，随着案件严重程度的增加，赔偿对量刑的影响相对下降，谅解对量刑的影响相对上升。<sup>[39]</sup>这种研究不仅用事实证明赔偿、谅解的实际作用，而且分别具体描述了它们各自的不同影响。而这种差异性影响是财富、身份因素解释不了的。

法教义学实证研究结果的相对性还体现在其建设性上。不同于单纯的摧毁、颠覆，对刑法现象实际范围、水平、程度的描述，为进一步的实践提供了建设性起点。例如，累犯从严是刑法中一个重要的量刑制度，但劳佳琦的实证研究发现，累犯前罪的轻重与性质、累犯前后罪的关系、累犯前罪刑罚执行完毕或赦免以后至再犯罪时间的长短，对量刑结果均不产生显著影响。据此，她提出，采用精算式评估有可能改进司法实践中人身危险性评估的质量。<sup>[40]</sup>同时，对策建议是描述性批判的自然结果。这里的批判不是消极证否，对策建议也不是凭空而来。描述性批判是积极的建设性批判，它不仅告诉人们何种理论站不住脚，还对实践该从何处再出发作出说明。在这方面，文姬的研究很有代表性。她基于精致的量化分析指出，在利用回归模型作为量刑模型预测刑罚量时，必须有较高的决定系数；在利用回归模型分析量刑影响因

[38] 参见白建军：《论不作为犯的法定性与相似性》，《中国法学》2012年第2期，第108页。

[39] 参见高通：《故意伤害案件中赔偿影响量刑的机制》，《法学研究》2020年第1期，第154页。

[40] 参见前引[32]，劳佳琦文，第247页。

素的相对重要性、计算基准刑时,则不需要有较高的决定系数。<sup>[41]</sup>这回答了此类研究长期以来的困惑。与口水战式的争辩相比,法教义学实证研究的结论更容易直接转化为可行的对策建议。

#### (六) 刑法教义学实证研究的类型

按照检验对象的不同,刑法教义学实证研究可以分为三种类型:(1)理论学说的教义学实证研究,其有助于刑法教义学理论研究提高自身信度以及对大范围法律实践的检验能力。(2)刑法规定的教义学实证研究,其有助于提高刑法修订以及刑事司法解释等的科学化水平。(3)司法实践的教义学实证研究,其有助于刑法教义学理论和规范的实效检验。

其一,刑法理论学说的教义学实证研究,其检验对象是一定的刑法理论学说;其事实材料可以是法条文本集合,也可以是司法裁判大样本,还可以是社会调查数据。“第一张骨牌”研究就证明,某种刑法理论二元思维模式中的任何一种具体理论,都不可能通吃所有刑法现象。刑法理论学说的教义学实证研究是刑法理论与刑法实践之间的桥梁。学界有观点认为,应把刑法第48条关于死刑适用条件的规定理解为,责任刑情节决定是否判处死刑,预防刑情节决定是否判处死刑立即执行。而王越基于九千多个故意杀人罪样本的实证分析,提出将死亡结果作为判处死刑的必要非充分条件,变减法机制为加法机制,建立严格的死刑从严情节体系和指导性的死刑从宽情节体系的对策建议。<sup>[42]</sup>这里的对策就是来自实践,经由理论批判又回到实践的对策。

刑法理论学说的实证检验,最诱人的课题之一,就是从德日引入的阶层犯罪论体系是否存在水土不服的问题。由于四要件理论不要求要件之间存在位阶顺序,其更像是一种犯罪特征的经验描述,同样一组案件事实更容易凑足入罪条件。相比之下,阶层论要求顺序操作的定罪程序,同样的评价对象更难通过这套程序,因而有更大的排除犯罪成立的可能性。但是,所谓“量刑反制定罪”现象的存在,<sup>[43]</sup>加之司法实践中大量变更罪名适用的客观现实,即使在中国推行阶层论,同样的事实证据满足指控罪名以外另一个罪名的定罪条件其实并不太难,从而可能软化中国司法实践中的阶层论。希望未来的某一天,阶层论的刑法教义学理论在中国司法实践中的实效检验能够成真。

其二,刑法规定的教义学实证研究,即基于一定法教义学理论及实证数据对某个或某些刑法规定进行正当性检验。对此,我曾尝试自上而下地对当时刑法规定的全部罪名进行分类、下降、赋权和量化分析,也曾自下而上地对分则罪名体系进行多元回归分析,目的都是试图发现罪刑关系体系的内在构造。<sup>[44]</sup>此外,按照美国《模范刑法典》的规定,构成犯罪的主观心理状态分为四种:有目的(purposeful)、明知(knowing)、鲁莽(reckless)、疏忽(negligent)。按照美国的诉讼制度,决定被告人是否有罪,需要由陪审团成员作出判断。因此,刑法规范是否被有效适用,在很大程度上取决于陪审团成员能否准确区分犯罪的四种主观心态。为此,美国学者招募志愿者进行了实验研究,结果发现,被试很难对某些界限作出区分,尤其是鲁莽与

[41] 参见文姬:《醉酒型危险驾驶罪量刑影响因素实证研究》,《法学研究》2016年第1期,第165页。

[42] 参见王越:《故意杀人罪死刑裁量机制的实证研究》,《法学研究》2017年第5期,第151页。

[43] 参见周建达:《“以刑定罪”的实践样态及其分析——以Y市法院的实证考察为基础》,《环球法律评论》2015年第1期,第20页。

[44] 参见白建军:《罪刑均衡实证研究》,法律出版社2004年版,第159页,第198页后插页;白建军:《再论罪刑均衡》,《新疆师范大学学报(哲学社会科学版)》2020年第1期。

明知的区别。由此不难推想，这种规定很可能会给司法适用造成混乱。所以，研究者提出，应该重新审视立法的明确性和立法者所期待的内容。<sup>[45]</sup>

对这类研究没必要顾忌实定法合理性假定。不仅前文对此已有阐释，学界实际上也有不少成功尝试。例如，梁根林等人近期发表多篇实证研究成果，论述死刑存废问题；<sup>[46]</sup> 欧阳本祺关于非法经营罪兜底性规定的实证研究发现，司法上的罪刑擅断应归因于立法上对罪刑法定的背离；<sup>[47]</sup> 毛玲玲关于金融犯罪司法状况的实证研究发现，应当限缩刑事立法的犯罪圈；<sup>[48]</sup> 江湖基于实证研究提出，应废除运输毒品罪的死刑，<sup>[49]</sup> 等等。在这个问题上，张明楷也不否定批判立法的可能，他认为实质的法益概念本身就具有批判立法的机能，即根据实质的法益概念废止缺乏保护法益的犯罪或者增设新罪。<sup>[50]</sup> 其实，绝对放弃对实定法的深入挖掘，才恰恰是对实定法合理性假定的曲解。

其三，刑事司法实践的教义学实证研究。这里所谓的司法实践不是具体个案，也不仅限于法院案例，而是某类案件法律适用各个流程环节中大量数据材料的总称。如前文所述，司法实践中的集体经验并非绝对正确，更不能直接拿来作为公正检验的标准。胡昌明基于一千余个盗窃罪样本的实证研究发现，当事人的身份因素对量刑轻重有显著影响，如地位越高，法度越宽。<sup>[51]</sup> 这显然暴露出了司法实践中存在的一些问题。

应当指出，目前此类研究多见于量刑实践的描述性批判，而少有犯罪论问题的实证检验，这应该与量刑问题研究比较适合各类量化分析工具的运用有关。不过，王华伟关于刑法但书规定的实证研究，就是一个典型的犯罪论司法实践的描述性批判。如何理解刑法第13条的但书规定，一直以来都有争议。于是，王华伟将120份以但书规定为出罪依据的裁判样本，放入以犯罪论体系中的基本要素为横坐标、以适用但书出罪的依据为纵坐标的分析框架中进行观察。结果发现，存在行为本该出罪但不应依据但书出罪的“错用”情形和行为本该出罪但仍然依据但书出罪的“滥用”情形，这说明司法实务部门对但书的内涵存在认识不足。<sup>[52]</sup> 崔志伟的实证研究也有类似发现，他认为但书规定在司法适用中常表现为“随性”、错用、滥用以及同案异判等粗疏化现象。<sup>[53]</sup>

实际上，具体案件的处理比法律制定更容易受到法律适用者自身认知、局限、偏好等主体性因素的影响，更容易出现这样或那样的“主体性外溢”问题。而有的实证研究对司法大数据中的问题与精华不加区分，盲目进行量化描述，不理解司法适用领域中正当性检验的重要

[45] See Matthew R. Ginther, Francis X. Shen, Richard J. Bonnie, Morris B. Hoffman, Owen D. Jones, Rene Marois, and Kenneth W. Simons, *The Language of mens rea*, 67 *Vanderbilt Law Review* 1327–1372 (2014).

[46] 参见前引〔31〕，梁根林等文，第1256页；梁根林、王华伟：《死刑替代措施的中国命运：观念、模式与实践》，《思想》2020年第5期，第57页；梁根林、马永强：《中国民众的死刑观念：观察、猜想与解析》，《清华法学》2020年第6期，第48页。

[47] 参见欧阳本祺：《对非法经营罪兜底性规定的实证研究》，《法学》2012年第7期，第119页。

[48] 参见毛玲玲：《金融领域刑事司法状况的实证考察与启示》，《法学》2014年第2期，第143页。

[49] 参见江湖：《律师对毒品犯罪死刑态度的实证研究》，《清华法学》2020年第5期，第59页。

[50] 参见张明楷：《论实质的法益概念——对法益概念的立法批判机能的肯定》，《法学家》2021年第1期，第80页。

[51] 参见胡昌明：《被告人身份差异对量刑的影响：基于1060份刑事判决的实证分析》，《清华法学》2018年第4期，第91页。

[52] 参见王华伟：《中国刑法第13条但书实证研究——基于120份判决书的理论反思》，《法学家》2015年第6期，第88页。

[53] 参见崔志伟：《“但书”出罪的学理争议、实证分析与教义学解构》，《中国刑事法杂志》2018年第2期，第3页。

性。解决此类问题的良药，正是司法实践的教义学实证研究。

除上述划分外，还可以按照具体内容不同，把教义学化的刑法知识分为犯罪论教义学、刑罚教义学和罪名（刑法各论）教义学三个部分。于是，两种划分相交织的结果，便有九种组合，形成一个九宫格体的刑法教义学实证研究框架。

犯罪论教义学理论学说的实证检验	犯罪论教义学刑法规定的实证检验	犯罪论教义学司法实践的实证检验
刑罚教义学理论学说的实证检验	刑罚教义学刑法规定的实证检验	刑罚教义学司法实践的实证检验
罪名教义学理论学说的实证检验	罪名教义学刑法规定的实证检验	罪名教义学司法实践的实证检验

表格中纵向三列分别由理论学说、刑法规定和司法实践组成，横向三行分别由犯罪论教义学、刑罚教义学和罪名教义学组成，由此形成的九个焦点便是九种研究类型。透过这个框架反观中国刑法研究，不难看出各部分的强弱并不均衡，所以，刑法教义学实证研究尚有许多值得进一步拓展和深化的空间。

---

**Abstract:** The differences between qualitative research and quantitative analysis and between normative science and factual science are only superficial differences between dogmatics of criminal law and empirical research of criminal law, which cover up the intrinsic values of these two. The rationality assumptions of positive law and that of the collective experience of empirical legal research are both relative. Therefore, there is no fundamental opposition between them and it is possible to study the dogmatics of criminal law with the method of empirical research. Dogmatics of criminal law and empirical research of criminal law are unified in practical rationality, meet in a large sample of criminal law cases, and intersect in the validity of law. Neither the empirical criminal law research without dogmatics nor the criminal law dogmatics without empirical research is conducive to the comprehensive development of criminal law research in China. The most extreme evil is the crime of illegally depriving another person's life and the most extreme negative social evaluation is the criminal punishment of depriving the offender's life. These two deprivations are the boundaries of all other deprivations. Therefore, the application of the death penalty in homicide cases is "the first piece of domino" in Chinese criminal law. Taking the criminal penalty dogmatics research of near 80,000 death penalty cases as an example, this paper shows that the empirical research of criminal dogmatics is a descriptive criticism of a large sample of criminal law phenomenon, including the dogmatic empirical research of criminal theories, of criminal law provisions and of criminal justice practices.

**Key Words:** dogmatics of criminal law, empirical research, the application of the death penalty, validity of law

---