中外法学 Peking University Law Journal
Vol.32, No.5(2020)pp.1211－1230

冲突与融合:认罪认罚从宽制度的本土化

魏晓娜*

摘 要 2019年底北京市两级法院判决的“余金平交通肇事案”揭示出认罪认罚从宽制度全面施行后法院、检察院之间的冲突。从表象上,这种冲突实质上是检、法两家对认罪认罚案件量刑主导权的争夺。其根源在于立法态度暧昧不明,没有明确区分两种不同的“从宽”逻辑。立法者之所以不肯明确承认量刑协商,是因为看到了“协商”背后隐藏的系统性风险。在以调查模式和层级模式为建构原则的中国刑事诉讼框架下,“协商”承载的是与之不相容的纠纷模式和同位模式的基本逻辑。因此,认罪认罚从宽制度存在进一步本土化的问题。对此有两种处理方案,一是管控冲突的烈度,重新定位认罪认罚从宽制度的性质与功能,使之实现从“案件处理机制”到“案件查明机制”的转型;二是管控冲突的范围,为“协商”施加适用范围上的限制。

关键词 认罪认罚从宽 量刑建议 量刑协商 诉讼模式

引 言

北京市两级法院于2019年底判决的“余金平交通肇事案”(以下简称“余金平案”),因承载了大量的问题点,引爆了刑事法学界关于认罪认罚从宽制度实施过程中诸多问题的激烈争论,成为2020年春季疫情初步缓解之后引发刑事法学界广泛讨论的一次“公共学术事件”。“余金平案”本身并不复杂。北京市门头沟区人民检察院指控,2019年6月5日21时许,被告人余

* 中国人民大学刑事法律研究中心教授。本文系国家社会科学基金项目“死刑限制的程序模式研究”(项目编号:15BFX73)和中国人民大学科研基金重大项目“人民陪审员制度改革研究”(项目编号:15XNL012)的阶段性成果。

金平酒后驾车,车辆前部右侧撞到被害人宋某致其死亡,撞人后余金平驾车逃逸。经北京市公安局门头沟分局交通支队认定,被告人余金平发生事故时系酒后驾车,且驾车逃逸,负事故全部责任。次日 5 时许,余金平到公安机关自动投案,供述自己的罪行。6 月 17 日,余金平家属赔偿被害人宋某的近亲属各项经济损失共计人民币 160 万元,获得被害人近亲属的谅解。余金平自愿认罪认罚,在辩护人的见证下签署具结书,并同意门头沟区人民检察院提出的有期徒刑三年、缓刑四年的量刑建议。2019 年 9 月 11 日,门头沟区人民法院以简易程序审结此案,认定了自首、初犯、赔偿损失、被害人家属谅解等法定、酌定量刑情节,但认为余金平身为纪检干部,酒后驾车,发生交通事故后逃逸,主观恶性较大,未采纳控方判三缓四的量刑建议,判处余金平有期徒刑二年。〔1〕一审判决作出后,余金平提出上诉,门头沟区人民检察院也提起抗诉。2019 年 12 月 30 日,二审法院北京市第一中级人民法院作出判决,认为余金平在明知发生交通事故且知道自己撞了人的情况下始终对这一关键事实不能如实陈述,因而不构成自首,一审法院认定具有自首情节并据此减轻处罚是错误的,最后改判余金平有期徒刑三年六个月。〔2〕

该案确实承载了认罪认罚从宽制度实施过程中诸多程序法和实体法上的争论点,诸如,“一审法院可否改变检察机关提出的量刑建议?”“检察机关对一审判决的抗诉是抗轻还是抗重?”“二审法院是否违反了上诉不加刑原则?”“余金平不承认撞人是否影响自首的构成?”“如果不构成自首,该案还是否属于认罪认罚案件?”等等。本文无意对控、辩、一审、二审各方立场分辨出个是非曲直,但此案的确提供了一个观察中国认罪认罚从宽制度的独特视角。余金平案之所以引人注目,爆点之一是该案中的法院、检察院一反中国刑事司法实践中人们习以为常的“你好我好大家好”的和谐局面,相互叫板,针锋相对,各不相让。门头沟区检察院提出判三缓四的量刑建议,门头沟区法院没有采纳,而是判处实刑二年有期徒刑;门头沟区检察院提起抗诉,北京市人民检察院第一分院支持抗诉,北京市第一中级人民法院反而加重判处有期徒刑三年六个月。有学者把余金平案中的检、法两家的角力形容为“神仙打架”,这确实是以往中国刑事司法实践中罕见的场面。余金平案的出现并非偶然,它是认罪认罚从宽制度全面施行之后诸多矛盾与冲突的一次集中爆发。本文关注的焦点在于,这些矛盾和冲突何以发生?未来又该如何从制度层面上进行破解?

一、冲突何以发生:三个层面的观察

(一)司法表象:量刑主导权之争

余金平案中两级法院、检察院的冲突与角力,其实质是检、法两家对量刑主导权的争夺。实际上,自 2018 年《刑事诉讼法(修正案)》生效以来,随着认罪认罚从宽制度的全面施行,检、法两家在量刑主导权方面的冲突逐渐成为公开的秘密。个别法官在具体案件中故意不采纳量

〔1〕 参见余金平交通肇事案,北京市门头沟区人民法院(2019)京刑初 138 号刑事判决书。

〔2〕 参见余金平交通肇事案,北京市第一中级人民法院(2019)京 01 刑终 628 号刑事判决书。

刑建议,甚至不通知公诉人调整量刑建议,径行判决,〔3〕诸如此类的现象,在认罪认罚司法实践中,并不鲜见。在最高人民检察院 2019 年 10 月 24 日召开的"准确适用认罪认罚从宽制度"新闻发布会上,最高人民检察院第一检察厅厅长苗生明也指出,目前检察机关提出的量刑建议,法院采纳率还不够高,实践中出现法官不采纳量刑建议,最终量刑与量刑建议之间相差一两个月甚至半个月的情形。〔4〕据统计,认罪认罚从宽制度入法之后,2019 年 1—5 月量刑建议法院采纳率仅为 51.75%,经检察机关采取积极措施,强力提升认罪认罚从宽制度适用率和量刑建议采纳率,量刑建议法院采纳率才逐步提升到 2019 年 1—6 月的 58%,1—9 月的 81.6%。〔5〕据最高人民检察检察长张军在 2020 年 5 月向全国人大做的工作报告,2019 年 12 月检察机关量刑建议法院采纳率又小幅回落到 79.8%。〔6〕

没有对比,就没有伤害。在认罪认罚从宽制度试点中期的 2017 年 12 月,检察机关量刑建议法院采纳率曾高达 92.1%,〔7〕到了试点末期的 2018 年 9 月底,量刑建议法院采纳率甚至一度高达 96.03%。〔8〕可见,在认罪认罚案件中,法院与检察院在量刑问题上的暗中角力,并非自始存在。2018 年新《刑事诉讼法》的正式实施,似乎是一个分水岭。

面对法院不采纳量刑建议的判决,一些检察官颇感委屈。《刑事诉讼法》第 201 条明确规定,对于认罪认罚案件,人民法院依法作出判决时,"一般应当"采纳人民检察院的量刑建议。即使不同意检察院的量刑建议,根据第 201 条第 2 款的规定,也应该先通知人民检察院调整量刑建议。当委屈转化为不平,有些检察机关对不采纳量刑建议的一审判决回应以抗诉。有些抗诉确实得到了二审法院的支持,如安徽省马鞍山市中级法院对一起开设赌场案作出的二审判决;〔9〕有的却再度遭遇二审法院的重击,如余金平案。

在法院看来,检察院提出精准量刑建议,〔10〕《刑事诉讼法》第 201 条又要求法院"一般应当"采纳检察院的量刑建议,相当于让法院自废武功,放弃量刑主导权。结合司法改革的大背景,检察系统内部 2018 年底开始落实内设机构改革,全面推行"捕诉合一",同时在司法责任制

〔3〕 例如,北京市第四中级人民法院判决的"张安故意伤害案",北京市第四中级人民法院(2019)京 04 刑终 8 号刑事判决书。

〔4〕 参见《最高检召开"准确适用认罪认罚从宽制度"新闻发布会》,载最高人民检察院官网,https://www.spp.gov.cn/spp/zgrmjcyxwfbh/zqsyrzrfckzd/index.shtml,最后访问日期:2020 年 6 月 2 日。

〔5〕 参见陈国庆:《认罪认罚从宽制度适用率逐步提升》,载最高人民检察院官网,https://www.spp.gov.cn/spp/zgrmjcyxwfbh/201910/t20191024_435923.shtml,最后访问日期:2020 年 6 月 2 日。

〔6〕 参见张军:《最高人民检察院工作报告(第十三届全国人民代表大会第三次会议)》,载最高人民检察院网,https://www.spp.gov.cn/spp/gzbg/202006/t20200601_463798.shtml,最后访问日期:2020 年 6 月 2 日。

〔7〕 参见周强:《关于在部分地区开展刑事案件认罪认罚从宽制度试点工作情况的中期报告》,载《人民法院报》2017 年 12 月 24 日,第 2 版。

〔8〕 参见胡云腾主编:《认罪认罚从宽制度的理解与适用》,人民法院出版社 2018 年版,第 278 页。

〔9〕 参见范跃红、徐静、陈乐乐:《认罪认罚了,量刑从宽建议为何未采纳》,载《检察日报》2019 年 9 月 22 日,第 1 版。

〔10〕 2019 年 10 月两高三部联合发布的《关于适用认罪认罚从宽制度的指导意见》第 33 条规定,"办理认罪认罚案件,人民检察院一般应当提出确定刑量刑建议"。2019 年 12 月修订公布的《人民检察院刑事诉讼规则》第 275 条规定,"量刑建议一般应当为确定刑"。

改革背景下,个体检察官在逮捕、起诉问题上掌握着相当的话语权,如果再在量刑问题上掌握主导权,权力过大,包含着极大的风险。〔11〕2019 年 5 月 20 日,《检察日报》又刊发社评,提出"在认罪认罚制度中,检察机关在诉讼中不仅是承上启下的枢纽和监督者,而且是罪案处理的实质影响者乃至决定者,具有主导作用、承担主导责任"。此后,"在认罪认罚案件中发挥主导作用"成为检察机关的惯常提法,〔12〕引起部分法官的抵触情绪,为检、法两家在量刑问题上的冲突埋下了伏笔。

(二)立法迷思:两种逻辑的交缠

在余金平案中,我们看到检、法两家在量刑主导权上的角力已经趋于白热化。这种现象之所以发生,与认罪认罚从宽问题上的立法思路不明直接相关。

众所周知,在试点初期,关于认罪认罚从宽制度的理解,就存在两种不同的思路。一种是"协商"的思路,认为认罪认罚从宽制度从本质上就是控辩协商。既然是协商,那么践行认罪认罚从宽制度的控辩双方之间就应当是平等关系,双方达成的关于量刑方面的一致意向在性质上应该是一种协议。按照这种逻辑,"从宽"是协商的结果。另一种是"政策实施"的思路,认为在中国的刑事司法中,认罪认罚从宽的政策和实践早已有之,所谓认罪认罚从宽制度,只不过是把以往的自首、坦白、立功、减刑、假释等制度系统化、正式化,"从宽"是对被告人通过"认罪认罚"表现出的社会危害性和人身危险性降低的一种制度性回应。按照这种逻辑,"从宽"不是协商的结果,而是立法者和司法者居高临下地给予已经认罪认罚的被告人的一种"恩惠"。上述两种思路的竞争和交缠并没有随着认罪认罚从宽制度的正式入法而终结,在新《刑事诉讼法》施行后,仍给研究者和司法者留下挥之不去的困惑。

一方面,为了与英美的辩诉交易划清界限,刑事诉讼法中凡涉及认罪认罚从宽制度的内容,都尽量回避使用"协商""协议""合意"等容易让人联想到"辩诉交易"的字眼。例如《刑事诉讼法》第 173 条第 2 款关于听取犯罪嫌疑人及其辩护人或者值班律师在相关事项上的意见的规定,被认为是最具有"协商"意味的立法表述,但自始至终未出现"协商"一词。〔13〕最高法院杨立新法官曾撰文指出:"《关于认罪认罚从宽制度改革试点方案》(以下简称《试点方案》)以及《试点办法》均回避使用'协商'一词,而是通过规定审查起诉阶段检察机关应就涉嫌犯罪事实、定罪、法律适用、从宽处罚的建议以及程序适用等问题听取辩方意见的方式,为控辩双方

〔11〕 时任最高法院大法官胡云腾就指出,"若强求一个集批捕、起诉权于一人的独办检察官在起诉时就提精准量刑建议,不仅勉为其难,而且权力过大,容易出问题"。胡云腾:《正确把握认罪认罚从宽 保证严格公正高效司法》,载《人民法院报》2019 年 10 月 24 日,第 5 版。

〔12〕 例如,胡莲芳、陈杨林:《如何推进检察机关落实认罪认罚从宽制度主导责任》,载《检察日报》2019 年 9 月 22 日,第 3 版。

〔13〕 《刑事诉讼法》第 173 条第 2 款规定:"犯罪嫌疑人认罪认罚的,人民检察院应当告知其享有的诉讼权利和认罪认罚的法律规定,听取犯罪嫌疑人、辩护人或者值班律师、被害人及其诉讼代理人对下列事项的意见,并记录在案:(一)涉嫌的犯罪事实、罪名及适用的法律规定;(二)从轻、减轻或者免除处罚等从宽处罚的建议;(三)认罪认罚后案件审理适用的程序;(四)其他需要听取意见的事项。"

的量刑协商提供程序保障”。〔14〕由此可见,立法回避使用“协商”的措词也是有意为之。一部国家立法,对于控辩之间的协商“欲说还休”,以这种“犹抱琵琶半遮面”的方式表达自身的真实立法意图,足见立法者在这一问题上的纠结心态。

又如,假若控辩双方就认罪认罚从宽达成一致意向,所形成的书面文件不是“协议”,而是“具结书”。根据《刑事诉讼法》第174条的规定,犯罪嫌疑人自愿认罪,同意量刑建议和程序适用的,应当在辩护人或者值班律师在场的情况下签署具结书。然而,从汉语词义来看,“具结书”是犯罪嫌疑人单方面向办案机关呈交的保证书,〔15〕只能约束签署具结书的犯罪嫌疑人,无论是法律上还是道义上,都不具有约束双方的性质和效力。这样的设计,与“协商”的旨趣相去甚远。

同样,控辩双方就量刑问题达成的一致意向并不落实为量刑协议,而是量刑建议。虽然认罪认罚案件中的量刑建议被解释为控辩协商合意的结果,〔16〕与非认罪案件中的量刑建议有本质区别。然而,也有学者对上述解释提出质疑,认为“犯罪嫌疑人认罪认罚是检察机关听取当事人双方意见的前提,而非通过控辩协商反过来促使犯罪嫌疑人认罪认罚,可见,该条款既难以解释为‘认罪协商’,也难以解释为‘量刑协商’”。〔17〕显然,就制度基本面来看,量刑协议与量刑建议,二者的旨趣大为不同:量刑协议主要面向控辩双方之间的关系,量刑建议主要牵涉检察机关与法院的关系。《刑事诉讼法》第201条第2款关于检察机关调整量刑建议的规定,遵循的实际上仍然是“量刑建议”的逻辑。

另一方面,立法又不自主地按照“协商”或者“协议”的逻辑设计相关法律规范。《刑事诉讼法》第201条规定,除该条明确列举的情形外,“对于认罪认罚案件,人民法院依法作出判决时,一般应当采纳人民检察院指控的罪名和量刑建议”。立法者在这里用了“一般应当”这样颇具争议的情态词,表达了对于法院尊重、顺应控辩双方意愿的殷切期待。这里的潜台词是,既然此时的量刑建议包含了控辩双方之间的合意,那么法院就不要轻易自行改变,否则,如果法院屡屡推翻控辩之间就量刑问题达成的一致意向,那么认罪认罚从宽制度就难以推行下去。殊不知,这种家长式的立法安排,已经超越了量刑建议所能够容纳的制度逻辑,立法要求法院“一般应当”采纳检察机关的量刑建议,实际上是自行“代入”了量刑协议的逻辑。因为,依照诉讼原理,量刑是法律适用问题,属于法官的职权范围,检察机关提出的量刑建议对法庭没有约束力,对审判中查明的犯罪行为如何量刑,属于法庭固有的职权。

然而,在认罪认罚从宽制度的适用范围问题上,立法又切换回“政策实施”的逻辑。目前立法在认罪认罚从宽制度适用案件范围和适用阶段问题上,漫无限制。只要具备认罪认罚的程

〔14〕 参见杨立新:“认罪认罚从宽制度理解与适用”,《国家检察官学院学报》2019年第1期,第59页。

〔15〕 从词源上来看,具结是指“旧时对官署提出表示负责的文字”。《辞海》(第6版),上海辞书出版社2009年版,第1181页。

〔16〕 参见陈国庆:《适用认罪认罚从宽制度的若干问题(上)》,载《法制日报》2019年11月27日,第9版。

〔17〕 熊秋红:“比较法视野下的认罪认罚从宽制度——兼论刑事诉讼第四范式”,《比较法研究》2019年第5期,第89页。

序适用条件，无论案件的严重程度，也无论是审判阶段、审查起诉阶段、侦查阶段，均可以适用。其背后的思想实际上是基于刑法中关于自首、立功、坦白等从宽处理的规定，这些制度的适用不受案件性质和严重程度的任何影响。问题是，一旦基于“政策实施”的制度逻辑不限制认罪认罚从宽的适用范围，那么基于协商逻辑的认罪认罚从宽也会自动扩大适用于几乎所有的案件。可以想象一下这会产生什么样的实践效果：无论何种性质、多么严重的案件，只要被告人认罪认罚，检察机关提出量刑建议，立法就要求法院“一般应当”采纳量刑建议。

所以，我们看到，刑事诉讼法对认罪认罚从宽制度的表述，一方面竭力避免使用“协商”“协议”“合意”等与“协商”关联度较高的措词，以与辩诉交易划清界限；另一方面又有意无意地在法条结构中植入“协商”的逻辑。然而在适用范围问题上，认罪认罚从宽制度又切换成“政策实施”的逻辑。立法态度纠结至此，司法实践中的混乱和纷争几乎势不可免。

（三）深层矛盾：“协商”背后的系统性冲突

立法者的态度为何如此纠结？那是因为看到了“协商”二字背后涌动的系统性冲突。

为了理解“协商”一词对中国刑事司法带来的冲击，本文借用比较刑事诉讼研究中一些常用的理论分析工具，力图多角度展现中国刑事司法的特性。在比较法学者看来，世界主要西方国家的刑事诉讼，可以归入不同的二元理论模型进行观察，例如中国刑事诉讼法学界所熟知的对抗制和审问制（adversarial system v. inquisitorial system），或者当事人主义与职权主义（adversary system v. inquisitorial system）的二元模型。然而，对抗制和审问制，或者当事人主义与职权主义的二元模型有自身的局限性，这种分类方法主要着眼于刑事诉讼程序（特别是审判程序）的开展和推进方式，对于刑事诉讼的基本性质、刑事诉讼中的各个决策主体（主要指法官、检察官、警察等刑事诉讼程序的主导者）的组织特性关注不多，因而在观察各个决策主体的行为时视角单一，解释力有限。为全方位理解中国刑事诉讼中各个决策主体的行为，本文再引入纠纷模式和调查模式、层级模式和同位模式两对理论模型，相当于增加了两个观察视角。当然，在西方学术语境中，这两对理论模型，与对抗制和审问制或者当事人主义与职权主义的理论模型一样，主要是以西方国家刑事诉讼实践为研究范本，比如，英美国家的刑事司法更接近对抗制（当事人主义）、纠纷模式和同位模式，欧洲大陆国家的刑事司法更接近审问制（职权主义）、调查模式和层级模式。然而，这并不影响上述理论模型在中国刑事司法问题上的解释力，尤其是纠纷模式和调查模式、层级模式和同位模式两种理论模型，可以让我们超越审判程序推进方式上的浅层次特征，关注更深层次的价值取向和决策主体的组织结构等特征。

1. 纠纷模式和调查模式〔18〕

纠纷模式和调查模式代表着对刑事诉讼性质和目的的两种不同理解。纠纷模式把刑事诉讼理解为控辩双方在相对消极的裁判者面前解决纠纷的活动，相应地刑事诉讼的基本目的也

〔18〕 纠纷模式与调查模式的英文为“the model of dispute”和“the model of official investigation”。这种区分至少在 19 世纪已经出现，其提出者已不可考。See Maximo Langer, “From Legal Transplants to Legal Translations: The Globalization of Plea Bargaining and the Americanization Thesis in Criminal Procedure”, *Harvard International Law Journal*, Vol.45, No.1, 2004, p. 20.

在于解决纠纷;调查模式则把刑事诉讼设计为司法官员依职权对案件真相进行调查的活动,刑事诉讼的基本目标被设定为查明真相。英美的刑事诉讼基本上可以归入纠纷模式,欧洲大陆国家大体上可以归入调查模式。刑事诉讼是致力于解决纠纷,还是查明真相,这种差异具有根本性的意义:它形塑了两种不同的程序展开和推进方式,即对抗式和审问式,或者当事人主义与职权主义模式;形成两套不同的话语系统;决定两种不同的权责配置方案;决定了对司法主体作用和角色的认知和期待,包括社会公众和司法主体对自身角色的理解。对抗式和审问式、当事人主义与职权主义,为人们常用且熟知,本文不再专门加以论述,这里仅解释后三个方面的影响。

(1)话语体系。两种不同的模式形成了不同的话语系统,即使相同的术语,两种话语系统也可能对之赋予不同的含义。以"真相"一词为例。在纠纷模式下,如果对案件真相产生争议,也有对真相的查明活动。然而,纠纷模式对真相的理解更具有相对性、合意性。如果控辩双方对案件事实达成了一致意见,无论是通过交易达成的一致还是通过约定达成的一致,那么,查明案件实际上是如何发生的,就没有那么重要了。[19] 然而,在调查模式下,对"真相"的理解更具有绝对的意味。无论控辩双方对于案件事实是否达成一致的认识,司法官员(包括法官)都被期待通过客观的调查查明实际上发生的案件真相。又如对"检察官"的理解。在纠纷模式下,检察官被理解为对诉讼结果存有利害关系的一方当事人;在调查模式下,检察官被定位为公正客观的国家司法官员,其职责在于调查案件真相,因此,检察官不能像当事人那样一味追求胜诉,必要的时候,可以为被告人的利益提起上(抗)诉。

有些术语,仅存在于一种话语体系,在另一种话语系统中是缺失的。例如"答辩"一词,仅存在于纠纷模式(或对抗式诉讼),一方当事人在正式场合(例如在英美的传讯程序中)对对方诉讼主张的正式承认,就构成一个"答辩"。一个合法有效的答辩,具有终结审判的法律效果。当然,在纠纷模式下也存在供述(自白)的概念,它是非在正式场合作出的认罪,只是一种证据,不构成答辩,不具有终结审判的程序法效力。在调查模式下,控辩双方的诉讼态度是不能左右司法主体探明真相的活动的,因而,在调查模式的话语体系中,"答辩"的概念是缺失的。在调查模式下,被告人对犯罪的承认,只能构成供述,供述只具有证据法意义,并不足以确立关于案件事实的"真相","真相"仍需要法官去认定。

(2)权责配置。调查模式之"调查",实际上指官方(法院、检察院、警察)主导的调查。[20] 为了便于司法主体查明案件真相,立法者为他们配置了广泛的程序权力,他们可以依职权主动采取调查行动,尽一切合法可能的手段查明案件事实。这一点,既可以理解为司法主体的权力,也可以理解为职责。德国《刑事诉讼法》第160条规定:"一旦检察院通过告发或其他途径获知犯罪行为嫌疑,应当对案情进行调查,以决定是否提起公诉。检察院不仅应当侦查对被指控人不利的情况,还应当侦查对其有利的情况,并且负责收集有丧失之虞的证据。检察院的侦查亦应涵盖对确定犯罪行为的后果具有意义的情况。……"第244条第2款规定了审判法院

〔19〕 Ibid., p. 10.

〔20〕 Langer, supra note 18, p. 23.

的职责:“为了调查事实真相,法院应当依职权将证据调查延伸到所以对于裁判具有意义的事实和证据上。”〔21〕从其文字表述可以看出,法院不受“当事人双方”向他提出的主张与证据的约束。其结果,所有的证据都是法院的证据,不“属于”任何一方当事人。法院负有查明案件事实的责任也意味着控辩双方没有提供证据的责任,检察官没有积极举证,并不必然导致败诉的后果,证明责任只在纯客观的意义上存在着。

在纠纷模式下,刑事诉讼被理解成控辩双方发生的一场纠纷,诉讼程序和证据也是按照这种观念来塑造。控辩双方都会在审前阶段开展各自独立的调查,所收集的证据也会自动分成“控方证据”和“辩方证据”。在法庭上,主要依赖于控辩双方的攻防活动将程序引向深入。法官被塑造成消极的裁断者,只被要求对控辩双方向他提出的争议事项作出裁决。

(3)对司法角色的认知和期待。对刑事诉讼性质的不同理解,不仅塑造了不同的诉讼结构和诉讼实践,而且内化为对司法主体的社会认知和自我认知。在纠纷模式下,控辩双方被塑造成程序的积极推动者,法官则是一个消极的裁断者,这不仅是社会对法官角色的认知和期待,也是法官对自身角色的理解和期待。一个过于积极地介入举证活动的法官总是要面对丧失公正立场的质疑。而在调查模式下,法官被认为负有澄清案件事实真相的义务,如果法官在没有穷尽一切合法可能的调查手段的情况下直接作出判决,属于“调查未尽”,可以构成上诉第三审之法定事由。〔22〕这种对于法官作用的定位和理解不仅为一般社会公众所接受——就社会公众而言,一个好的法官一定是积极查明案件真相,不为控辩双方的任何一方所蒙蔽和左右的法官,这进而也成为法官个体对自身角色的认知和理解。上述认知和理解,是通过长期的诸如法学教育、司法培训、立法塑造,以及与其他法律主体的互动等一系列社会化的过程而规训和习得。

对司法角色的社会期待和(司法者)自我认知当然是与立法互动的结果,法官或积极或消极的司法角色最初是由立法所塑造,但日积月累,潜移默化,这种对法官角色的定位和期待就逐渐内化为司法文化的有机组成部分。然而,立法的转型并不会立竿见影地带来角色认知的变化。比较法学家观察到一个有趣的对比:意大利于 1988 年实现了刑事诉讼的全面转型,从以往的职权主义转向英美的对抗制。作为过渡,意大利《刑事诉讼法》第 507 条仍保留了法官依职权主动调查证据的规定,即法官在审判过程中仍可自行传唤证人并对证人发问。实际上,美国《联邦刑事诉讼规则》第 614 条也有类似的规定。有趣的是,两个相似的条文在美国和意大利司法实践中的命运却截然不同。《联邦刑事诉讼规则》第 614 条在美国没有产生多大影响,实践中法官甚少动用这一权力,因为一个积极寻找真相的法官显然与对抗制之下法官应当扮演的消极角色相矛盾。然而在意大利,第 507 条却成为司法实践中最常被法官援引的条文。〔23〕原因在于,在以往的大陆法传统之下,法官被认为对裁判的准确性负有责任,当审判

〔21〕 宗玉琨译注:《德国刑事诉讼法典》,知识产权出版社 2013 年版,第 155、193 页。

〔22〕 参见林钰雄:《刑事诉讼法(上)》,元照出版有限公司 2013 年版,第 75 页。

〔23〕 William T. Pizzi & Mariangela Montagna, “The Battle to Establish An Adversarial Trial System in Italy”, *Michigan Journal of International Law*, Vol.25, No.2, 2004, p. 447.

的主导权已经由法官手中转移到控辩双方手中,法官们一时很难适应这种消极的角色。不仅法官如此,其他诉讼角色一时也很难适应这种急剧的变化。例如,如果检察官由于疏忽而没有传唤对证明某一犯罪要素非常重要的证人,法官会认为他必须传唤这个证人,而不能眼睁睁看着案件由于检察官的疏忽而流产。在这种情况下,美国的法官可能更愿意让控方的起诉失败。同样,在意大利,如果法官认为一项证据可能有利于被告人而辩方没有出示,法官传唤证人的压力尤其强大,因为意大利的法官被视为负有家长式地保护被告人的责任,如果辩护律师不够老练,法官会认为自己应该传唤本该由辩护律师传唤的证人,或者帮助辩护律师询问证人。某一对辩方有利的证人应该传唤律师却没有申请传唤,意大利的法官们更难以坐视一个可能无辜的人被定罪。〔24〕

2.层级模式与同位模式〔25〕

达马斯卡教授提出的层级模式与同位模式主要关注刑事诉讼中决策主体(法官、检察官、警察)的组织特性。其中层级模式比较贴合欧洲大陆的刑事司法实践,同位模式则更接近英美的刑事司法状况。

(1)层级模式。层级模式以判决的确定性为主要目标,"同案同判"的正义原则特别受到推崇;判决的确定性要求政策的统一性;在政策统一性诉求的驱动下,自然会导致权力组织形态的集中化;在集中化的权力组织形态下,各个权力主体并非自主的个体——权力只是委托给他们行使,他们行使权力时必须接受严密的监督;决策权也因此走向严密的层级化。〔26〕

警察部队和检察机关的主要组织原则是集中化,主要的行动规则是统一性。为了保证起诉政策的一致性,检察官享有的起诉裁量权十分有限,越是在严重的案件中越是如此。就法官而言,各国宪法为法官提供的独立性保障会产生一定的离心倾向。然而在欧洲大陆的刑事司法中,早已发展出抵御这种离心趋势的秘密武器——全面的上诉审查。在现代国家萌芽时期,欧洲大陆的统治者就通过建立层级化的官僚体系控制之前封建割据的地区。下级法官做出的判决要接受更接近权力中枢的上级法官的审查,这种上诉机制特别适合加强中央集权的政治需要。因此,在欧洲大陆,上诉审从一开始就被设计为对案件进行全面、彻底审查的机制,至少在初次上诉中是如此。这意味着不仅要对法律错误,而且要对事实认定、刑罚适用等方面进行全方位的审查。当事人的上诉是一种受到鼓励的行为,因为只有借助于上诉,才能够实现对下级法官审判活动的监督。因此,"上诉权"被上升为一种宪法权利,上诉被设计为对当事人而言低成本、无风险的程序。也正是由于对刑事司法的层级控制无处不在,所有的官方活动、所有重要事项的处理都必须记录下来,保存在案卷中,全程留痕以备上级部门的审查,所以官方的文件和案卷具有特别的重要性。

---

〔24〕 Ibid., pp. 447－449.

〔25〕 英文为 The Hiararchical Model and the Coordinate Model,也有学者将 The Hiararchical Model 翻译为"阶层模式"。

〔26〕 Mirjan Damaska, "Structures of Authority and Comparative Criminal Procedure", *Yale Law Journal*, Vol.84, No.3, 1975, p. 483.

(2)同位模式。同位模式以作出最适合于每一个案件具体情况的判决作为优先价值目标。判决确定性当然也是一个重要的价值,但其分量远不如在层级模式之下,为了判决的确定性和统一性而牺牲具体案件的最优解决方案是不能接受的。刑事审判中由于陪审团的使用,判决结果具有相当大的不确定性,因此"同案同判"的正义原则在英美法中并不像在欧洲大陆那样受到重视。在同位模式下,正义总是个别化的正义,只有对案件具体情形有着充分的了解,才可能做出最适于个案的判决。同位模式下,理想的判决主体一定是单层的,而不是有很多层级。高级别的法官在选择适合于案件的最优解决方案方面,不仅不具有优势,相反由于时间、空间的局限,还经常具有一些劣势。因此,即便是在整个司法系统处于基座位置的裁判者,在他自己所主导的领域,往往也享有最高的权威。

在同位模式下,政策的统一性固然也是一个重要的价值,但很难奢求通过一个个独立的决策主体来自动实现。〔27〕英美的警察组织具有浓厚的地方性色彩,在很多重要的方面都是离散化的。检察机关同样也是如此,尤其是在美国,大多数州的检察官都是地方选举产生的官员,实际上没有上级,享有极其广泛的自由裁量权,只要他认为能够更好地为自己的选民服务,他在起诉问题上作出的任何决定实际上是无法审查的。普通法所承认的陪审团废止权(nullification)是另一个极端例证。陪审团废止权意味着,即使面对无可辩驳的证据,即使有法官明确的指示,陪审团仍可以作出无罪裁决,从而拒绝适用实体法。普通法对这一特权提供的解释是,陪审员必须运用当地的正义观来作判决,使刑法规定适应案件的具体情形。〔28〕由于陪审团作出裁决的时候不需要解释判决理由,无论是有罪裁决还是无罪裁决,要对这种交织着事实问题和法律问题的笼统性裁决进行欧洲大陆那样的上诉审查,实际上是不现实的。

3.中国的刑事司法模式

如前所述,虽然西方比较法学家在提出纠纷模式和调查模式、同位模式与层级模式的理论模型时,其研究对象未必包括中国,然而,我们在两对模型的对比中,很容易找到中国刑事司法的位置,甚至可以毫无违和感地将中国的刑事司法直接代入调查模式和层级模式的逻辑。

中国的刑事司法可以直接代入调查模式。刑事诉讼从整体思路上,被设计为由不同阶段首尾相继而构成的调查程序,警察、检察机关、法院分别是主导着各个阶段的调查活动。同时,在中国的刑事诉讼法中,对案件事实真相的执著追求贯穿着刑事诉讼的始终。"准确、及时地查明案件事实"被确定为刑事诉讼法的基本任务,"以事实为依据,以法律为准绳"是刑事诉讼基本原则之一,"真相"一词甚至被写作具有哲学上真理意味的"真象"。〔29〕这种具有绝对意味的真相观,没有给任何的妥协、协商、合意留下共生的空间。同样,中国刑事诉讼法的话语体系也与调查模式高度一致。刑事诉讼法中只有"供述"的概念,没有"答辩"的概念,不会因为被告人对控方事实主张的认可而直接认定犯罪事实成立。相反,"只有被告人供述,没有其他证

〔27〕 Ibid.

〔28〕 Damaska, supra note 26, p. 512.

〔29〕 参见《刑事诉讼法》第2条、第6条。《刑事诉讼法》第53条规定:"公安机关提请批准逮捕书、人民检察院起诉书、人民法院判决书,必须忠实于事实真象。故意隐瞒事实真象的,应当追究责任"。

据的,不得认定被告人有罪和处以刑罚”。检察官不是纠纷的一方当事人,而是负有调查责任的中立、客观的司法人员。在司法主体的权责配置上,一方面,为保证查明案件真相,刑事诉讼法赋予各司法主体尤其是法官较为广泛的权力,法官甚至享有“庭外调查权”。〔30〕另一方面,司法主体也承担较为广泛的责任,不限于诉讼法上的责任,即判决事实不清可能面临上级法院撤销原判、发回重审的否定性评价,〔31〕而且还涉及职业责任,即本着“让审理者裁判,由裁判者负责”的原则,让法官对其履行审判职责的行为承担责任,甚至对办案质量终身负责。而案件被上级法院发回重审、改判,均属于办案质量的负面指标。〔32〕在中国传统文化中,理想的法官形象是“包青天”式的明察秋毫的法官。在当代,在全面而广泛的司法责任压力下,法官不可能将自身简单定位为纠纷解决者,更不能接受“橡皮图章”式的角色。

中国的刑事司法也与层级模式相契合。在中国这样中央集权的单一制国家,法律适用的统一性被赋予特别重要的地位。而层级模式下的全面的上诉审查,特别适合这种价值取向。上诉不仅是对当事人的救济,而且是上级法院对下级法院行使审判监督的重要途径,为了利用当事人的上诉带动上诉审查机制发挥作用,上诉被设计为低成本、低风险的行为,上诉不需要说明理由,实行上诉不加刑原则,以免除被告人的后顾之忧。法院分层级,审判也分层级,而且级别越高,权威越大。上级法院可以在事实认定、法律适用等任何方面否定下级法院的判决。判决的一致性和政策的统一性被赋予了重要的价值。例如,余金平案的二审判决书将“关于一审法院对余金平判处实刑是否属于同案不同判问题”作出一个重要的问题点专门进行回应和解释,〔33〕也有研究特意对余金平交通肇事案之类案进行大数据分析来判断余案的量刑是否适当,〔34〕由此可以管窥“同案同判”的正义观在中国是何等地根深蒂固。

可见,中国的刑事司法不仅在核心价值追求上,而且话语体系、权责配置、角色认知以及上诉制度设计方面,都更契合调查模式和层级模式。

回到本节开首的问题,“协商”的观念从根本上与调查模式是不相容的。在调查模式下,“真相”只能是查明的,不可能是协商、妥协或者协调出来的。更重要的是,“协商”本质上承载

〔30〕《刑事诉讼法》第196条规定:“法庭审理过程中,合议庭对证据有疑问的,可以宣布休庭,对证据进行调查核实。人民法院调查核实证据,可以进行勘验、检查、查封、扣押、鉴定和查询、冻结”。

〔31〕根据《刑事诉讼法》第236条的规定,第一审判决事实不清楚或者证据不足的,第二审法院可以在查清事实后改判;也可以裁定撤销原判,发回原审人民法院重新审判。

〔32〕《最高人民法院关于进一步全面落实司法责任制的实施意见(法发〔2018〕23号)》第2条提出,要全面落实司法责任制,坚持“让审理者裁判,由裁判者负责”。《最高人民法院关于完善人民法院司法责任制的若干意见(法发〔2015〕13号)》第25条规定:“法官应当对其履行审判职责的行为承担责任,在职责范围内对办案质量终身负责。”《最高人民法院关于进一步全面落实司法责任制的实施意见(法发〔2018〕23号)》第15条规定:“强化案件质量评查。坚持案件常规随机评查、重点评查、专项评查相结合,重点评查发回重审案件、改判案件、信访案件以及曾纳入长期未结、久押不决督办范围的案件”。

〔33〕参见余金平交通肇事案第(二)部分“关于抗辩争议焦点的综合评述”中第3点,北京市第一中级人民法院(2019)京01刑终628号刑事判决书。

〔34〕参见小包公团队:《余金平交通肇事案之类案大数据分析》,载微信公众号“小包公”,2020年4月21日上传。

着纠纷模式的逻辑。所以,在中国的刑事诉讼法中接纳“协商”一词,就相当于在调查模式中植入了纠纷模式的逻辑,必然引起“排异反应”,检法两家在量刑问题上的激烈冲突,就是这种排异反应的表现之一。排异的结果无非两种:一是夹带着纠纷模式“私货”的协商机制充当了“特洛伊木马”,在原本由调查模式占据的领域攻城略地,最终将刑事诉讼改造成符合纠纷模式逻辑的样态。这不仅意味着刑事诉讼要放弃对真相的执著追求,放弃原有的话语体系,更要对原有的决策主体权责配置作出根本性改变,这对于当前中国的刑事诉讼而言,基本上是不可能的事情。对司法角色的社会认知和自我认知的改变,属于法文化层面的变化,更是难上加难。第二种结果,在话语体系、权责配置、角色认知无法撼动的前提下,体现纠纷模式逻辑的“协商”的生存空间被挤压到十分稀薄的程度,甚至没有“协商”,只有“合意”。刑事诉讼立法以“具结书”代替协议书,以“量刑建议”代替量刑协议,甚至连合意的形式要素都不齐备。立法基调如此,实践中的协商意味就更为稀薄了,常见的做法是,检察官收到辩护律师的法律意见后,自行提出量刑建议,然后让犯罪嫌疑人和律师签字确认。〔35〕

(四)小结

综上所述,以余金平案为代表的认罪认罚从宽制度实施过程中的各种矛盾和冲突,肇源于两个方面因素的叠加:一是认罪认罚从宽制度中植入的“协商”逻辑与中国刑事司法原有的价值定位、诉讼模式及其配套制度必然发生碰撞与抵牾;二是认罪认罚从宽制度自身的设计没有清晰地区分两种从宽的逻辑,没有廓清两种逻辑的适用领域,特别是没有限定“协商”的边界。这不仅进一步放大了认罪认罚从宽制度与原有诉讼模式的冲突,而且使得适宜开展量刑协商的案件由于重重顾虑而不能充分展开,导致司法实践中“协商”的因素极为稀薄,最终损害认罪认罚被告人的利益。这意味着,认罪认罚从宽制度目前确实存在进一步本土化的问题。同时,为了充分发挥认罪认罚从宽的制度效能,也需要给控辩双方之间的“协商”以应有的生存空间。

二、从冲突到融合:两种方案

要从制度层面解决,至少是缓解认罪认罚从宽制度与中国原有诉讼制度之间的矛盾和冲突,必须有的放矢,围绕上述两个方面的问题探索具体方案。相应地,也会存在两种解决问题的思路:一是管控冲突的烈度,避免认罪认罚从宽制度与原有诉讼模式的正面冲突,对认罪认罚从宽的性质和功能进行重新界定,该方案可以最大限度保留目前立法设计的认罪认罚从宽制度框架;二是管控冲突的范围,接受认罪认罚从宽制度与现有诉讼模式之间的冲突,但把冲突管控在较小的范围。这意味着,给基于“协商”逻辑的认罪认罚从宽制度划定适当的范围。当然,其前提是,必须对基于两种不同逻辑的“从宽”进行明确区分。上述两种方案,既可以叠加使用,也可以择一适用。选择任何一种,都可以在相当大的程度上避免或者缓解认罪认罚从宽制度与原有诉讼制度的正面冲突。然而,叠加使用的后果,可能会在很大程度上压制认罪认

〔35〕 参见关振海:“检察机关落实认罪认罚从宽制度的四个建议”,载《检察日报》2019年8月2日,第3版。

罚从宽制度效能的发挥。因此,本文更倾向于择一适用。以下分述之。

(一)冲突烈度之管控:认罪认罚从宽制度性质之修正

“协商”观念与刑事诉讼的调查模式之间的冲突,并非中国所独有。毫无疑问,作为调查模式典型代表的欧洲大陆国家在引入协商机制时也遇到过类似的问题。为了绕过各种制度障碍,各国刑事司法中的协商机制并非美式辩诉交易的忠实拷贝,而是进行了脱胎换骨式的改造。〔36〕在这方面,以德国最为典型。了解德国化解调查模式与协商机制之冲突的手法,或许能为中国破解当前认罪认罚从宽制度困局提供借鉴。

在德国,刑事诉讼中的协商机制包括起诉协商、判决协商和处罚令协商,〔37〕起诉协商和处罚令协商都规避了正式审判,相比之下,判决协商更有可能与原有的奠定调查模式基本权力结构的条款发生正面冲突。〔38〕

德国是通过修改《刑事诉讼法》,增设新的第257c条引入判决协商的。然而,为避免上述冲突,当初起草修正案的联邦司法部一开始就明确了协商制度的立法宗旨:“立法规制的目的在于为协商制度的具体运作提供详细的规范指导,但不会对个案中必要的自由裁量权加以限制”。为此,草案规定了如下基本原则:“①量刑的基本原则不变。即量刑必须根据被告人的刑事责任进行;②刑事诉讼法的基本原则不变。协商并不是法院作出判决的基础和前提,法院仍将致力于案件事实真相的发现;③最大限度的透明化。协商只能在审判程序中进行,对于审判程序之外的协商,法院必须公开告知,协商必须全文记录并在判决中指出;④法律救济上没有任何限制。放弃上诉的承诺并不是进行协商的前提。协商后的判决仍将接受全面的审查。被

〔36〕 意大利特别程序的多样化以及辩诉交易程序的变形,主要是为了规避强制起诉原则。在意大利,强制起诉规定于《宪法》第112条,对辩诉交易制度的构建形成刚性的制度障碍。为此,意大利的辩诉交易程序只允许对量刑进行交易,而不能对罪名进行交易。

〔37〕 1975年,德国在废除预审法官的同时,在规定起诉法定原则的德国《刑事诉讼法》第152条之后,增设两个例外,即第153条和第153a条。第153条赋予检察官在轻微案件中无条件撤销案件的权力,第153a条赋予检察官附条件撤销案件的权力。在不涉及公共利益的轻微案件中,如果检察官认为继续侦查会占用太多时间,可以向辩方表示,如果嫌疑人同意履行一定的条件,如向慈善机构或者国家缴纳一定数额的金钱,检察官就不再提起公诉。根据第153a条的规定,适用这一程序需要有法官、检察官和被告人的一致同意。此项改革,催生出起诉协商。处罚令协商来源于德国的处罚令程序。在程序的开始,检察官要准备一份处罚令草案,详细说明案件情况和申请的刑罚,刑罚往往仅限于罚金和交通案件中的吊销驾照。然后,检察官要把处罚令草案和案卷一并交给法官。法官一般只是例行公事地签署处罚令。之后,处罚令以挂号信的形式寄给被告人。对被告人而言,接受处罚令意味着通过支付罚金和承认有罪避开公开审判带来的尴尬、时间和名誉损失。对检察官和法官而言,处罚令是处理案件的高效工具。因此,控辩双方围绕处罚令产生大量的协商。所谓判决协商,根据2009年修改后的德国《刑事诉讼法》第257c条,主要是指法院和诉讼参与人(包括法官、检察官、被告人、辩护人、附诉人)对诉讼进程和结果进行的协商。参见魏晓娜:《背叛程序正义:协商性刑事司法研究》,法律出版社2014年版,第51-52、60页。

〔38〕 这样的条款包括,德国《刑事诉讼法》第155条第2款:“在此(起诉所称的犯罪行为和所指控的人员)范围内,法院有权且有义务独立活动;尤其在刑法的适用上,不受提起的控告拘束”。第244条第2款:“为查清真相,法院依职权应当将证据调查涵盖所有对裁判具有意义的事实和证据材料”。宗玉琨译注,见前注〔21〕,第158-159、193页。

告人应被充分告知协商的相关情况。”〔39〕上述原则,大部分为新增设的第 257c 条所吸收,〔40〕成为立法的要求。该条第 1 款明确指出,“第 244 条第 2 款的规定不受影响。”“刑事诉讼法的基本原则不变……法院仍将致力于案件事实真相的发现”,这意味着协商机制的引入并没有冲击到“调查模式”的基本价值追求和权力配置框架;“法律救济上没有任何限制”,则表明德国仍秉持“层级模式”下的一贯立场。那么,德国是如何做到的?

根据第 257c 条,协商的前提和重要内容是被告人作出供述(因此德国的判决协商又被称为“供述协商”)。但是,法院并不因此而被免除查明案件事实真相的义务,法院依然有义务依职权调查事实真相,直到形成对被指控人定罪的内心确信。〔41〕可见,协商的目的不是为了获取“认罪答辩”,只是为了取得被告人的当庭“供述”。有了供述,审判还是要继续进行,然而有了供述之后,更有利于法官查明案件真相,认定案件事实,由此可以大幅度缩短审判的持续时间,而这恰恰成为催生“协商”的强大动力。在供述协商过程中,法官依然是德国刑事司法中最活跃的角色,并没有变成消极的裁判者,更没有变成“橡皮图章”。因此,在德国供述协商实践中,法官延续了其在刑事诉讼中一贯的积极、强有力的形象,只有他能够保证量刑承诺的最终实现,检察官的撤消起诉在很大程度上也要受他节制,案件的任何处理方案都不可能绕过他。

德国处理协商制度与原有诉讼模式之间冲突的秘诀在于,协商得来的供述协议并不直接决定案件的最终处理,因此并非处理案件的机制,供述协议仅仅充当了发现真相的工具。经过这样的改造,供述协议不仅没有成为“调查模式”的对立面,反而为后者所吸纳。为了获得被告人详实的口供,以“从轻”或“减轻”情节的形式给被告人提供量刑方面的优惠,这在大陆法系从来都不是什么新鲜事儿。所以,协商制度的引入并未颠覆德国原有的诉讼模式及其相适应的权责分配框架,立法者为了使协商制度适应原有的调查模式,放弃了在纠纷模式下“案件处理机制”的定位,将协商制度纳入事实查明机制,较为妥善地化解了协商制度的引入对原有诉讼模式带来的冲击。

〔39〕 参见黄河:“德国刑事诉讼中的协商制度浅析”,《环球法律评论》2010 年第 1 期,第 125 页。

〔40〕 德国《刑事诉讼法》第 257c 条[法院与诉讼参与人之间的协议]规定:“1.适宜的情形下,法院可以依据下列规定与诉讼参与人就程序的进一步发展和程序的结果进行协商。第 244 条第 2 款的规定不受影响。2.协议的标的限于:能够构成判决及其所属裁定之内容的法律后果,案件查明程序中其他的程序相关的措施,以及诉讼参与人的诉讼行为。任一协商都应当含有认罪内容。有罪宣告和矫正及保安处分不得作为协议的标的。3.法院告知协议可能包含的内容。其可以基于案件的所有情况及综合量刑考量,自由裁量给出刑罚的上限和下限。诉讼参与人有机会提出意见。如果被告人和检察院同意法院提出的建议,则协议成立。4.如果忽视或者出现新的法律上或者事实上的关键情况,且法院确信原先的量刑幅度与行为或者罪责不匹配的,则法院不受该协议的拘束。如果被告人采取的进一步诉讼行为与法院预测所依据的行为不一致,此规定同样适用。这些情形下,被告人的认罪不予适用。法院应当就此背离毫不迟疑地进行通知。5.应当告知被告人法院依据第 4 款背离预测结果的前提和条件”。参见岳礼玲、林静译:《德国刑事诉讼法典》,中国检察出版社 2016 年版,第 110 页。

〔41〕 参见李倩:“德国认罪协商制度的历史嬗变和当代发展”,《比较法研究》2020 年第 2 期,第 93 页。

回到中国,随着认罪认罚从宽制度的施行,协商元素的实际引入,〔42〕不可避免地与原有的诉讼模式及其相关的基本价值取向、权责配置、角色设定发生冲突。中国刑事诉讼致力于追求案件事实真相和法律统一适用的基本价值取向不可能发生动摇,这是中国刑事诉讼制度的立足之本,也是刑事诉讼基本原则、权责配置等基本制度得以确立的基石。因此,为避免包含有协商逻辑的认罪认罚从宽制度与原有的诉讼框架发生冲突,必须明确界定认罪认罚从宽的性质和功能。从上述德国的经验可以看出,只有将认罪认罚从宽定性为案件查明机制,而非案件处理机制,才能避免二者发生正面冲突。

然而,目前《刑事诉讼法》第201条要求法院"一般应当"采纳检察机关提出的量刑建议,实际上是在"案件处理机制"的功能定位下做出的立法安排。这必然与调查模式之下对法官的权责配置、角色定位正面对撞。须知中国的法官仍负有查明案件真相的责任,立法为此给法官配置了较为广泛的调查权,同时法官也对判决的准确性负有不可推卸的责任。这不仅仅是诉讼法上的责任(判决事实不清可能被上级法院改判,或者撤销原判、发回重审),而且是一种司法责任(法官要对其履行审判职责的行为承担责任,甚至对办案质量终身负责,案件被上级法院发回重审、改判,也会在案件质量评查中受到的否定性评价)。〔43〕立法要求法官"一般应当"采纳检察机关的量刑建议,同时司法体制却要求法官背负如此沉重的责任,权责配置明显失衡。不仅如此,这种立法安排也打破了法、检之间的权力平衡。近年来,随着检察机关内设机构改革的推进,"捕诉合一"改革落地,检察机关内部权力下沉,检察官集审查批捕、审查起诉权力于一体。只要犯罪嫌疑人认罪认罚,那么无论什么样的案件,检察官听取犯罪嫌疑人意见后,即可提出确定刑量刑建议,然后立法就要求法院"一般应当"采纳检察机关提出的量刑建议。这种立法设计,实际上是将审查批捕权、审查起诉权与实质量刑权集于检察官一身。而法官的权力,不仅要受到体制内各种力量(审判委员会、上级法院等)的牵制,还要面临检察机关的蚕食,检、法权力配比明显失衡。凡此种种,皆肇源于立法者实质上是在"案件处理机制"的定位下作出的制度安排。要改变这种局面,认罪认罚从宽制度需要回归"案件查明机制"的性质和功能定位,这意味着一系列制度安排也需要随之作出调整:

首先,目前《刑事诉讼法》第201条关于人民法院"一般应当"采纳人民检察院量刑建议的规定,应当从立法上删除。这种表述,不仅不符合语言规范,逻辑上自相矛盾,而且不符合诉讼法理,在主要法治国家均没有先例。实际上,即使是体现控辩双方共同意愿的量刑协议,也并不必然约束法官。以美国为例,美国《联邦刑事诉讼规则》第11条(c)款规定了两种类型的量刑交易,

〔42〕《刑事诉讼法》虽没有出现"协商"一词,但认罪认罚从宽制度施行过程中官方也不再讳言"协商",比如2019年10月24日两高三部联合发布的《关于适用认罪认罚从宽制度的指导意见》第33条要求检察机关在提出量刑建议前,尽量与辩方"协商一致"。另外,法、检系统的代表性人物在讨论认罪认罚从宽制度时,也直接使用"协商"一词。参见胡云腾,见前注〔11〕;陈国庆,见前注〔16〕;苗生明、周颖:"认罪认罚从宽制度适用的基本问题",《中国刑事法杂志》2019年第6期,第3—29页。

〔43〕《最高人民法院关于进一步全面落实司法责任制的实施意见(法发〔2018〕23号)》第15条规定:"强化案件质量评查。坚持案件常规随机评查、重点评查、专项评查相结合,重点评查发回重审案件、改判案件、信访案件以及曾纳入长期未结、久押不决督办范围的案件"。

一种是完全不能约束法官的量刑协议，而且法官会告知被告人，即使法官没有采纳量刑建议，被告人也不能撤回答辩；〔44〕另一种是“有约束力的答辩”(binding plea)，控辩双方在被告人答辩有罪后可以确定量刑，而且，一旦法庭接受了该协议，就必须按照协议中确定的刑罚量刑。〔45〕然而，第二种量刑交易在司法实践中非常少见，原因是，许多法官拒绝接受这样的协议，认为这种协议“不被许可地”(impermissible)侵犯了法官的量刑权。〔46〕正常情况下，控辩双方达成的关于量刑的一致意向，法官没有特别的理由一般都会采纳，这是一种事实状态。但是，如果将这种状态变成立法要求，不仅理论上说不通，实际效果还会适得其反。目前我国司法实践中存在的个别法官对检察机关提出的量刑建议本着“主刑加减 1 个月、罚金增减 1000 元、缓刑考验期增减 1 个月”〔47〕的原则进行微调，故意不采纳量刑建议的现象，就是明证。

其次，检察机关提出量刑建议应以幅度刑量刑建议为主，确定刑量刑建议应当慎提。目前立法对量刑建议应当是幅度刑还是确定刑量刑建议没有明确规定。两高三部《关于适用认罪认罚从宽制度的指导意见》第 33 条和《人民检察院刑事诉讼规则》第 275 条都规定量刑建议“一般应当”是确定刑量刑建议。之所以推崇确定刑量刑建议，主要着眼于控辩量刑协商的角度，精准的量刑建议可以提高协商条件的明确性和协商结果的可预测性，有助于被告人协商地位的改善，整体上对被告人有利。然而，一旦着眼于检、法关系，学界和实务界对于此种立场都存在不同意见。例如，黄京平认为，幅度刑量刑建议是使司法建议权与司法裁定权恰当协调的最优方式。〔48〕胡云腾认为，如果是合议庭审判的案件，最好提有幅度的量刑建议，因为这类案件往往事实情节较多，不易权衡。有幅度的量刑建议既能体现对量刑的慎重，也能体现对合议庭的尊重。〔49〕德国在这一问题上的立场比较坚决，不允许在法庭审理之外提出一个“精准的刑罚”，只能提出有上限和下限的幅度刑。〔50〕本文认为，从理顺检、法关系计议，幅度刑量刑建议更为合理。至于由此给被告人带来的可能的期待利益方面的损失，可以通过构建下述程序性保障来解决。

---

〔44〕 美国《联邦刑事诉讼规则》第 11 条(c)(1)(B)规定：“(答辩协议可以指明控方将)建议或者同意不反对辩方申请特定的量刑或者量刑幅度，或者量刑指南某一条款、政策陈述或者量刑情节，适用或者不适用(该建议或要求不约束法庭)”。接着第 11 条(c)(3)(B)规定了这种量刑协议的效力：“第 11 条(c)(1)(B)中规定的答辩协议类型，法庭必须告知被告人，如果法庭没有听从该建议或者申请，被告人无权撤回答辩”。

〔45〕 美国《联邦刑事诉讼规则》第 11 条(c)(1)(C)规定：“(答辩协议可以指明控方将)同意，具体的量刑或者量刑幅度，或者量刑指南某一条款、政策陈述或者量刑情节，适用或者不适用(该建议或要求不约束法庭)”。接着第 11 条(c)(3)(A)规定了这种量刑协议的效力：“法庭可以接受、拒绝该协议，或者推迟到法庭审查量刑前报告之后再作判决”。

〔46〕 See Vanessa A. Edkins & Allison D. Redlich, *A System of Pleas*, New York: Oxford University Press, 2019, p. 13.

〔47〕 参见闵丰锦：“检察主导抑或审判中心：认罪认罚从宽制度中的权力冲突与交融”，《法学家》2020 年第 5 期，第 108 页。

〔48〕 参见黄京平：“幅度刑量刑建议的相对合理性——《刑事诉讼法》第 201 条的刑法意涵”，《法学杂志》2020 年第 6 期，第 100 页。

〔49〕 参见胡云腾，见前注〔11〕。

〔50〕 参见李倩，见前注〔41〕，第 93 页；另见宗玉琨译注，见前注〔21〕，第 204 页，注①。

再次,完善告知程序。目前刑事诉讼法在认罪认罚从宽制度适用的不同阶段均规定有告知程序,但告知的内容非常笼统,即犯罪嫌疑人、被告人的诉讼权利和认罪认罚从宽的法律规定。〔51〕相比之下,德国要求告知的内容更为丰富且具有针对性。德国《刑事诉讼法》第257c条确立了两个层面上的告知义务:一是抽象层面的,即该条第5款要求法院"应当告知被告人,法院依据第4款背离承诺结果的前提条件和后果"。也就是说,法院必须事先告知被告人,在什么情况下他不会接受之前量刑协议的约束,可能会对被告人判处更重的刑罚。2018年,德国联邦法院在判决中强调,法院的"告知义务"对供述协商制度具有重要意义;"告知义务"必须在供述协商实现之前作出;告知义务的目的在于确保被指控人了解潜伏在供述协商过程中以及被指控人作出认罪供述的风险;〔52〕二是具体层面的,即该条第4款规定的,如果法院决定改变协议中的量刑,"应当不迟延地告知将背离承诺"。也就是说,当法官在具体案件中决定要改变原来协商好的量刑之前,应当告知被告人他将不按原先的承诺去量刑。未来我国完善告知程序时可以加以借鉴,使得告知的内容更为丰富和具体。

又次,确立法院不接受量刑建议时的救济,保护被告人在量刑协商中的期待利益。从确定刑量刑建议转向幅度刑量刑建议,被告人的期待利益本已受损,如果法院又未采纳检察机关依据双方协商提出的量刑建议,会造成被告人的期待利益遭受更为实质性的损害。为了维持程序的公平性和对等性,如果被告人为争取宽大量刑,基于对控辩之间协议的信赖而作出认罪陈述,而法院最后决定不采纳检察机关依据协议提出的量刑建议,那么作为弥补,被告人在协商中所作的供述原则上也不能作为定罪量刑的证据使用。同时,在法院不采纳量刑建议时排除被告人依量刑协议作出的认罪供述的可采性,反过来还可以对法院形成制约,防止法院轻率地拒绝检察机关依量刑协议提出的量刑建议。

最后,保障被告人的上诉权,不得以放弃上诉权作为量刑协商的前提条件。在层级模式下,上诉制度的功能不仅限于救济被告人,而且具有监督功能。〔53〕在德国,协商放弃上诉权在立法和联邦法院的判决中都是不允许的。〔54〕在法国,根据庭前认罪程序做出判决,并不影响控辩双方行使上诉权。〔55〕在我国认罪认罚从宽制度施行的初期,保留这种监督性上诉尤其必要。认罪认罚从宽制度本身与中国原有的诉讼模式和诉讼原则就存在冲突和抵牾,立法者不可能对所有问题和风险都事先作出预测并提前防范,只能随着该制度在全国范围内的大规模施行而逐渐暴露,充分呈现,如果允许以放弃上诉权作为协商的前提条件,大量的问题可

〔51〕 参见《刑事诉讼法》第120条、第173条、第190条。

〔52〕 参见李倩,见前注〔41〕,第97页。

〔53〕 我国第二审程序贯彻全面审查原则,即是明证。《刑事诉讼法》第233条规定:"第二审人民法院应当就第一审判决认定的事实和适用法律进行全面审查,不受上诉或者抗诉范围的限制。共同犯罪的案件只有部分被告人上诉的,应当对全案进行审查,一并处理"。

〔54〕 参见李倩,见前注〔41〕,第98页。

〔55〕 法国《刑事诉讼法》第495-11条第2款规定:"对院长或者院长委派的法官作出的裁定,被判刑人均可按照第498条、第500条、第502条的规定,向上诉法院提起上诉;检察院得按照相同条件提起附带抗诉"。参见《法国刑事诉讼法典》,罗结珍译,中国法制出版社2006年版,第316页。

能就此被掩盖,那么认罪认罚从宽制度也失去了进一步发展、完善的动力和活力。保留被告人的上诉权,不仅可以强化被告人的协商地位,而且始终保留上级法院监督审查的可能性,更有利于认罪认罚从宽制度的健康发展,也为未来制度层面的进一步完善提供了可能。

(二)冲突范围之管控:限制协商机制的适用范围

第二种应对方案是对协商机制的适用施加明确的范围限制,在此范围内,立法明确承认控辩之间的量刑协商,并按照协商的逻辑作出相应的制度安排。这意味着,立法应当区分认罪认罚从宽的两种不同逻辑:一是控辩协商的逻辑,此种逻辑需要局限在一定的案件范围;二是政策实施的逻辑,超越协商机制适用范围的被告人认罪认罚,应当基于"宽严相济"刑事政策,根据刑法中关于"自首""立功""如实陈述"等量刑情节的处理原则给予一定的宽大处理。按照协商逻辑作出制度安排当然会对原有的诉讼模式产生一定冲击,但由于协商逻辑的适用范围受到限制,这种冲击可以被管控在一定的范围内,不会对刑事诉讼基本原则造成伤筋动骨式的影响。在协商逻辑下的制度安排,可以不再考虑是否与调查模式和层级模式发生冲突。以下是对这种协商机制初步设想:

1.案件范围

除德国外,在具有调查模式和层级模式特征的国家和地区,在刑事诉讼中引入协商机制时,为了减缓对原有诉讼模式的冲击,一般都会将严重案件排除出协商程序的适用范围,即使不能完全消除冲突,有可以有效地将冲突管控在一定的范围。例如,被认为最接近美国辩诉交易的意大利"依当事人请求判处刑罚"程序,适用该程序判处的刑罚有明确的限制,即提供 1/3 的量刑减扣后判处的刑罚不得超过 5 年,这意味着该程序适用于可能判处 7.5 年监禁刑以下刑罚的案件。〔56〕在法国,庭前认罪程序仅适用于主刑为罚金或者 5 年以下监禁刑的轻罪,在检察官提议执行监禁刑时,刑期不得超过 1 年,也不得超过当处监禁刑期的一半。〔57〕我国台湾地区 2004 年引入的协商程序,其适用范围受到两个方面的限制:一是案件范围,该程序适用于"除所犯为死刑、无期徒刑、最轻本刑三年以上有期徒刑之罪或高等法院管辖第一审案件者外"的案件;〔58〕二是科刑范围,"法院为协商判决所科之刑,以宣告缓刑、二年以下有期徒刑、拘役或罚金为限"。〔59〕那么,我们应当将协商程序限制在什么样的案件范围?2020 年《最高人民检察院工作报告》披露,1999 年至 2019 年,被判处三年有期徒刑以上刑罚的占比从 45.4%降至 21.3%。这意味着,目前有接近 80%的刑事案件在三年有期徒刑以下量刑。同时,简

〔56〕 参见意大利《刑事诉讼法》第 444 条。该条最初的规定是,在综合考虑全案情形和第 444 条提供的 1/3 的量刑减扣后,最终判处的刑罚不得超过 2 年。为了拓宽该量刑协商机制的适用范围,2003 年第 134 号法律将上述限制放宽到 5 年。这意味着将依当事人请求适用刑罚程序的适用范围从可能判处 3 年监禁刑以下刑罚的案件,扩大到可能判处 7.5 年监禁刑以下刑罚的案件。See Mitja Gialuz, "The Italian Code of Criminal Procedure: A Reading Guide", in Mitja Gialuz, Luca Luparia & Federica Scarpa (eds.), *The Italian Code of Criminal Procedure: Critical Essays and English Translation*, Stampato: Wolters Kluwer, 2014, p. 45.

〔57〕 参见法国《刑事诉讼法》第 495-7 条和第 495-8 条第 2 款,见前注〔56〕,第 313-314 页。

〔58〕 参见我国台湾地区"刑事诉讼法"第 455-2 条第 1 款。

〔59〕 参见我国台湾地区"刑事诉讼法"第 455-4 条第 2 款。

易程序也以三年有期徒刑作为程序适用的分界线,〔60〕因此,本文认为协商程序的适用范围确定为“可能判处三年有期徒刑以下刑罚”的案件较为适宜。

2.量刑协商

被告人的认罪,可以是协商的前提,也可以是协商的结果。关于协商的事项,我国台湾地区“刑事诉讼法”第455—2条第1款既允许“量刑协商”,也允许“负担协商”,包括:“一、被告愿受科刑之范围或愿意接受缓刑之宣告。二、被告向被害人道歉。三、被告支付相当数额之赔偿金。四、被告向公库或指定之公益团体、地方自治团体支付一定之金额。”德国可以协商的事项范围更为广泛,包括:①法院有权采取的措施,例如批准检察官不起诉决定和调查收集证据的决定;②被告人有权做出的行为,例如放弃进一步调查取证的申请、作出供述以及承诺对被害人进行赔偿;③检察官和附带起诉人有权做出的行为,例如放弃诉讼继续进行的申请。〔61〕我们可以立足于《刑法》总则第三章的规定,博采众长,允许就以下事项进行协商:①被告人可被判处的刑罚或者刑罚幅度;②《刑法》第37条和第37条之一规定的非刑罚性处置措施和《刑法修正案(九)》增设的从业禁止措施;③《刑法》第36条规定的由犯罪行为导致的经济损失的民事赔偿责任;④特定的诉讼行为,例如变更强制措施。控辩双方经协商达成一致意见的,应当签署量刑协议。然后,检察机关提起公诉,并根据已经达成的量刑协议附具明确的量刑建议。量刑建议可以是确定刑量刑建议,也可以是幅度刑量刑建议,取决于与被告人协商的结果。

3.法院的审核、裁判与救济

对于控辩双方通过协商达成的量刑协议,法官应着重审核以下三个方面:一是犯罪事实的真实性;二是检察官所提议的刑罚的适当性,即所提议之刑罚是否与犯罪情节、被告人的人身危险性相适合;三是程序的运作是否合乎程序要求,例如,律师是否在场、检察官是否履行了告知义务,以及被告人认罪是否明确、自愿、非出于外在压力等等。被告人应当出庭。法官应当告知被告人享有的法定权利以及因适用协商程序而丧失的权利。如果法官确认被告人自愿认罪、放弃诉讼权利,而且认定犯罪事实有充分的证据支持,协商的刑罚适当,程序合法,则可以在控辩双方协商的范围内做出判决。否则,则恢复普通程序或者简易程序进行审理,被告人作为协商结果的认罪陈述不得在日后的审判中作为证据使用。协商判决作出后,原则上不得提起上诉,但被告人的认罪、协商行为不具有自愿性,或者案件不属于协商程序适用范围的除外。

三、结　语

认罪认罚从宽制度正式入法、全面推行迄今尚不足两年,其实施效果如何,制度设计有无

〔60〕 我国《刑事诉讼法》第216条规定:“适用简易程序审理案件,对可能判处三年有期徒刑以下刑罚的,可以组成合议庭进行审判,也可以由审判员一人独任审判;对可能判处的有期徒刑超过三年的,应当组成合议庭进行审判”。第220条规定:“适用简易程序审理案件,人民法院应当在受理后二十日以内审结;对可能判处的有期徒刑超过三年的,可以延长至一个半月”。根据此两条,可能判处的有期徒刑超过三年与否,决定着简易程序的审判组织和审判期限。

〔61〕 参见德国《刑事诉讼法》第257条c。

偏误，目前仍处于观察、试错的阶段。“余金平交通肇事案”反映出检察院、法院之间的角力和冲突，检、法两家对于认罪认罚案件中量刑主导权的争夺只是最表象的原因。检、法两家之所以如此激烈地争夺量刑主导权，问题仍然出在立法上。立法态度不明朗，不肯明确承认“协商”，没有给控辩协商提供充分的制度空间。在制度安排上，两种不同的“从宽”逻辑——“协商”的逻辑和“政策实施”的逻辑没有清晰的边界，相互交缠，相互取代，成为司法实践中许多混乱和纷争的根源。立法者之所以在“协商”问题上态度如此纠结，根本原因是看到了“协商”观念背后隐藏的系统性风险。“协商”体现的是纠纷模式和同位模式的基本逻辑，而中国的刑事司法在基本价值追求、话语体系、司法角色的社会认知和自我认知等方面均体现着与前者不相容的调查模式和层级模式的基本逻辑。二者的冲突势不能免。因此，认罪认罚从宽制度在中国目前仍面临进一步本土化的问题。为充分发挥控辩协商的制度效能，同时避免协商制度与我国原有诉讼模式的正面冲突，本文提供了两种方案：一是重新界定认罪认罚从宽制度的性质和功能，使之从“案件处理机制”转化为“案件查明机制”，并据此对相关制度进行调整。该种方案一方面可以最大限度地保留目前立法所确定的认罪认罚从宽制度的大框架，同时也可以避免与中国原有诉讼模式的正面冲突；二是接受这种正面冲突，但把冲突管控在有限的范围，在此范围内，可以最大限度地按照协商的逻辑作出制度安排。

Abstract: Yu Jinping Case judged by two courts in Beijing in late 2019 revealed the conflict between the courts and the procuratorates after the full enforcement of the system of lenience on admission of guilt and acceptance of punishment. On the surface, the conflict comes from their fight for dominant authority in sentencing. But the fight for sentencing authority actually originates from the equivocal attitude of the legislature. Two different logics are not differentiated: the logic of negotiation and policy enforcing. Why the legislature declined to give a decent position to negotiation between the prosecution and the defense, is because they have seen the systematic risk behind negotiation. Against the basic framework of Chinese criminal justice characteristic of official investigation model and hierarchical model, negotiation reflects the incompatible logic of model of dispute and model of coordinate. The conflict between them has to be solved. There are two options: first, to redefine the nature and function of system of leniency on admitting guilt and accepting punishment, and to transform it from a case solving mechanism to a fact-finding mechanism; second, to control the scope of the conflict, and to impose limits on the applying scope of negotiation mechanism. But within the applying scope, to follow the logic of negotiation to solve the cases where the defendants admit guilt and accept punishment.

Key Words: Leniency on Admitting Guilt and Accepting Punishment; Sentence Recommendation; Negotiation on Sentence; Procedure Models

（责任编辑：车 浩）