

数字时代失信惩戒法治化的新进阶： 从制裁失信到管控风险

陈国栋*

摘要：将失信惩戒纳入行政处罚体系是当前失信惩戒法治化的主流进阶。这无法解决将违约、违法行为视为失信行为并加以公法制裁的合法化难题，又不能充分救济相对人，也难以充分承担社会信用制度的功能期待，还有碍于社会信用制度的体系化。从管控资源配置风险出发，信用是基于信用数据的交易可信度评判工具，失信惩戒是为了管控资源配置风险，而非对失信行为进行法律制裁。实施以风险管控为目的的失信惩戒，既符合行政机关的资源配置主体身份，也符合社会治理创新的需要与大数据时代社会治理革新的趋势。在大数据时代，失信惩戒法治化框架之建构当以信用算法的规制为中心，以风险管控原则统领信用算法，以个人信息权益与相应国家保障义务为具体抓手。

关键词：失信惩戒；法治化；风险管控；大数据；社会治理革新；信用算法

自2014年国务院出台《社会信用体系建设规划纲要（2014—2020年）》（国发〔2014〕21号）（以下简称“《纲要》”）以来，失信惩戒在社会治理领域迅速扩张。为解决失信惩戒“规范位阶过低、惩戒对象偏差、惩戒范围过大、手段与目标不合比例与程序控制缺乏”等问题，^[1]不少学者主张将其纳入行政处罚体系，由行政处罚法来规范。^[2]在修订行政处罚法的过程中，全国人大常委会部分委员也如此建议。^[3]2021年修订的行政处罚法以很大力度回应了这一呼声。不仅从业禁止、缔约资格排除等当下信用立法中的主流失信惩戒类型，被归为行政处罚中的资格罚，而且类似于黑名单的“通报批评”等声誉类惩戒，也被归为行政处罚中的名誉罚。有学者据此指出，在我国行政处罚法第2条为“行政处罚”增设定义条款后，“从实质层面将大部分失信惩戒措施定性为行政处罚已经争议不大”。^[4]当然，也有学者认为失信惩戒不是行政处罚，且坚持主张失信惩戒应当贯彻“过惩相当”原则。^[5]将“失信”界定为“过”，其实在失信惩戒的行为归属上与行政处罚论者并无本质区别。

* 大连海事大学法学院教授，法学博士。本文系国家社科基金重大项目“社会信用体系的法律保障机制研究”（项目编号：21&ZD199）、辽宁省社会科学规划基金重大委托项目“数字辽宁背景下数字政府的法治化研究”（项目编号：L22ZD002）的阶段性研究成果。

[1] 参见王瑞雪：《政府规制中的信用工具研究》，载《中国法学》2017年第4期，第159页；沈毅龙：《论失信的行政联合惩戒及其法律控制》，载《法学家》2019年第4期，第124页；沈岩：《社会信用体系建设的法治之道》，载《中国法学》2019年第5期，第34—41页。

[2] 参见秦前红、陈芳瑾：《地方信用立法的探索模式研究》，载《法治社会》2021年第4期，第89页；胡建森：《“黑名单”管理制度——行政机关实施“黑名单”是一种行政处罚》，载《人民法治》2017年第5期，第83页；王伟：《失信惩戒的类型化规制研究——兼论社会信用法的规则设计》，载《中州学刊》2019年第5期，第46页；张淑芳：《〈行政处罚法〉修订应拓展处罚种类》，载《法学》2021年第11期，第89页。

[3] 《全国人大常委会委员建议：将列入失信名单明确为行政处罚种类》，载中国网，<https://baijiahao.baidu.com/s?id=1689554274586660171&wfr=spider&for=pc>，最后访问日期：2023年1月8日。

[4] 谭冰霖：《处罚法定视野下失信惩戒的规范进阶》，载《法学》2022年第2期，第36页。

[5] 参见罗培新：《法理明晰 切中肯綮 有力助推营商环境优化》，载《中国信用》2021年第1期，第117页。

应该说,在公法学界普遍主张且这种主张已经为立法实务界逐渐接受的当下,将失信惩戒视为行政处罚进而法治化具有操作简单、成本低廉等优点。但是,失信惩戒起源于经济交易领域,是市场主体在收集各种信用数据从而评估交易风险的基础上采取的风险规避措施,是一种通过对交易风险的管控、预防而非强制性制裁来治理失信的治理手段。^[6]就此而言,将失信惩戒视为对失信行为的强制性法律制裁,背离了失信惩戒的本源与失信治理逻辑,且将失信惩戒改头换面为与其本源完全不同之物。^[7]这不仅凭空造就了难以解决的合法性难题及其他困境,还会使《纲要》所预期的失信治理效能落空,更因割裂失信惩戒与个人信用数据的有机关联而难以适应大数据时代社会治理的革新趋势与需要。^[8]欲解决上述困境,失信惩戒的法治化应当以管控风险而非制裁失信为进路。

一、将失信惩戒视为行政处罚的困境

尽管失信惩戒与行政处罚看起来有一定相似性,但将其视为行政处罚,会陷入如下难以突破的困境。

(一) 难以解决合法性问题

一是失信惩戒何以针对违约行为的问题。当下不少信用立法将违约行为视为失信行为,施加失信惩戒。^[9]但是,违约行为是私法行为,本应只由私法调整。将对其施加的失信惩戒视为行政处罚,就产生私法行为何以被认定为公法领域的违法行为并被施加公共制裁的难题,进而产生公权力侵犯私法自治、契约自由的问题。这个矛盾已有学者指出,为此,其将违约行为剔除出失信行为阵营。^[10]但是,这极大地限缩了社会信用制度的失信治理场域,背离了《纲要》“改善市场信用环境”的初衷,给失信惩戒带来了严重的合法性危机:若是不能治理违约这种市场领域公认的失信行为,失信惩戒的意义何在?为解决这一问题,有学者提出将诚信义务法律化,即将守信从道德义务上升为法律义务。^[11]这一进路被不少学者批评,认为其带来了法律的道德化。但问题不在于法律道德化,因为法律就是最低的道德,问题在于诚信这一美德难以法律化。道德法律化以道德具有公共性为前提,^[12]但守约之德并非公德,它不处理公共场域的道德关系,无需也不能上升为公域中的公德进而上升为法律义务。

二是何以将违法定性为失信进而施加公法制裁的难题。对违法行为施加作为行政处罚的失信惩戒,意味着违法行为除了违反诸如治安管理法律规范等特定公共管理法律规范外,还在失信意义上被当

[6] 早在21世纪初,林毅夫等经济学家就主张借鉴发达国家的通行做法,引入、建立覆盖全社会的信用奖惩制度,通过失信惩戒来解决市场交易领域失信泛滥的问题。后来一些经济学学者进一步探讨、界定了其风险管控而非法律制裁属性。参见王征宇:《美国的个人征信局及其服务》,中国方正出版社2003年版,第6页;林均跃:《社会信用体系原理》,中国方正出版社2003年版,第50—51页;王富全:《征信体系建设中的失信惩戒机制分析》,载《金融研究》2008年第5期,第187页;贺学会、王海峰、王小曼:《企业信用行为与失信惩戒机制:一个基于信用资本的分析框架》,载《金融研究》2008年第10期,第150页。

[7] 参见陈国栋:《私法信用惩戒的法理及其启示》,载《暨南学报(哲学社会科学版)》2021年第11期,第150页。

[8] 林均跃指出,失信惩戒是由数字驱动的机制,在社会信用体系的信用信息基础设施和信控服务业的共同支撑下运行。林均跃:《论公共和市场两种不同类型的失信惩戒机制及其互补关系》,载《征信》2022年第1期,第11页。一些法学者正是从信用治理与大数据的内在关联出发,主张从大数据时代数字治理而不是单纯的行政处罚的角度出发来理解失信惩戒及其法治化。参见戴昕:《理解社会信用体系建设的整体视角:法治分散、德治集中与规制强化》,载《中外法学》2019年第6期,第1472—1473页;刘晗、叶开儒:《平台视角中的社会信用治理及其法律规制》,载《法学论坛》2020年第2期,第62—63页;虞青松:《算法行政:社会信用体系治理范式及其法治化》,载《法学论坛》2020年第2期,第36—37页。

[9] 参见《上海市社会信用条例》第30条、《河南省社会信用条例》第32条、《辽宁省公共信用信息管理条例》第28条、《内蒙古自治区社会法人失信惩戒办法(试行)》第12条。

[10] 参见彭鐔:《失信联合惩戒制度的法治困境及出路——基于对41份中央级失信惩戒备忘录的分析》,载《法商研究》2021年第5期,第55页。

[11] 罗培新:《社会信用法:原理·规则·案例》,北京大学出版社2018年版,第59页。

[12] 参见孙海波:《道德立法的法哲学省思》,载《学术月刊》2021年第5期,第109页。

作一种独特、独立的法律失范(违法)行为。但违法何以失信?失信何以作为行政处罚的理由?这两个一脉相承的问题得不到解决,失信惩戒既难以作为“失信”惩戒区别于既有行政处罚,也难以解释其将“违法”视为“失信”加以治理的合法性与效能,进而难以解释行政机关对其管辖范围之外的违法行为施加联合惩戒的合法性。^[13]前一问题并没有受到行政处罚论者的关注。虽然有学者注意到了41份中央级失信联合惩戒合作备忘录所设定的失信惩戒基本上都有实定法依据,^[14]但没有注意并回应这一问题:这些早已为上位法规定的措施,何以在备忘录中被称为“失信”惩戒?有学者则从“法律是公民应当共同信守的契约”出发,认为违法行为亦属失信行为。^[15]这一理由显然难以勾连“失信”与“违法”。

(二) 难以满足公民权利救济需要

《纲要》提出,要使失信者“一处失信,处处受限”。为此,当下信用立法不仅设置了相应的黑名单制度,还明确要求其他公共机关鼓励市场主体实施联合惩戒,尽管该失信者可能并未违反其他行政管理法规,并未侵犯其他市场主体私权。联合惩戒由此成为失信惩戒的常用手段,失信者遭受几个甚至十几个行政机关与不计其数的市场主体的失信惩戒成为平常之事。这样一来,能否通过行政处罚框架来救济相对人,就不能不令人生疑。

根据韦伯所提炼的“暴力与强制只能由国家垄断,由国家行使”的现代行政法原则,^[16]作为强制性制裁的行政处罚只能由国家实施。据此,将失信惩戒视为行政处罚,无法相容于“行政处罚乃公权力制裁”这一行政处罚法教义学的根本法理。因为,当失信惩戒不得不依赖市场主体的声誉制裁才能产生制裁效果时,^[17]就意味着其制裁性取决于私人而非国家。这就难免导致失信惩戒过惩失当,违背行政处罚所应遵守的过罚相当原则与相应的明确性原则。一方面,在面向社会公布黑名单的情况下,相对人很可能陷入“网暴”式困境,^[18]因一个违法或违约行为而承受难以计数、难以觉察的失信惩戒,行为的失范性与后果的严重性难保相称;另一方面,因为联合惩戒主体不明确,惩戒的明确性无法实现:“有无制裁效果、制裁的程度无法由行政机关控制,而受制于种种社会因素。因而,这是一种可能违反法治主义(侵害的测定可能性和预见可能性较低)的手段。”^[19]就此而言,行政处罚框架难以有效保障相对人权益。有学者即指出,虽然我国新修订的行政处罚法将通报批评列入名誉罚,开辟了将列入黑名单、公布违法事实等失信惩戒措施纳入处罚框架的通道,但并不能实质性解决过惩失当、救济不足的问题。^[20]

(三) 难以实现《纲要》的失信治理预期

联合惩戒揭示了这样一个现实:没有市场主体的广泛参与,就没有有效的失信惩戒。有学者甚至指出,公共型失信惩戒机制“单打独斗”不是治理失信的终极办法,市场型失信惩戒机制才是根治失信的

[13] 有学者指出,违法不一定等于失信,因此对违法行为予以信用惩戒不一定有合法性。参见沈岿:《社会信用体系建设的法治之道》,载《中国法学》2019年第5期,第28页。

[14] 参见彭箴:《失信联合惩戒制度的法治困境及出路——基于对41份中央级失信惩戒备忘录的分析》,载《法商研究》2021年第5期,第56页。

[15] 罗培新:《社会信用法:原理·规则·案例》,北京大学出版社2018年版,第56页。

[16] [德] 马克斯·韦伯:《经济与社会》,林荣远译,商务印书馆2004年版,第731页。

[17] 如有学者认为,“如果将违法行为人的相关信息在特定范围内予以公布,将其行为纳入社会诚信体系记录中,让了解相关信息的单位或者个体对该行为人的信誉产生质疑,那么该行为人必然会对这样的处罚有所畏惧”。张淑芳:《〈行政处罚法〉修订应拓展处罚种类》,载《法学》2021年第11期,第89页。

[18] 即“与实施网暴的轻松和低成本相比,受害者遭受的羞辱不成比例地严重”。See Kristine L. Gallardo, Taming the Internet Pitchfork Mob: Online Public Shaming, the Viral Media Age, and the Communications Decency Act, 19 (3) Vanderbilt Journal of Entertainment & Technology Law 721, 729 (2017).

[19] 王贵松:《论行政处罚的制裁性》,载《法商研究》2020年第5期,第27页。

[20] 朱芒:《作为行政处罚一般种类的“通报批评”》,载《中国法学》2021年第2期,第164—165页。

厉害“后招”。^[21]一个新的问题由此产生:市场主体参与联合惩戒的动力是什么?这个问题非常重要。缺乏动力的失信惩戒机制不可能稳定地发挥“制裁”功能以实现社会信用制度的失信治理目标。而行政处罚思路在这方面也有碍于这一目的的实现。在将失信惩戒视为法律制裁思维的主导下,当下信用立法没有采纳经济学界对失信惩戒的界定,而是将市场主体实施的失信惩戒规定为“提高失信主体交易成本”之举,^[22]以避免突破法律制裁只能由国家垄断这一政法原则。但这一界定只会削弱市场主体参与失信惩戒的意愿,使立法者对失信惩戒的期待落空。

因为市场主体一般只关心自己的利润,“在商言商”地与所有人做生意,不会像有些学者所认为的那样,仅仅因为行政机关将相关违法、违约信息纳入诚信体系就实施声誉制裁。比如,食药监督管理部门根据《关于实施餐饮服务食品安全监督量化分级管理工作的指导意见》(国食药监食〔2012〕5号),用“大笑”“微笑”和“平脸”三种卡通形象在餐饮服务单位显著位置显示餐饮服务食品安全等级,但消费者往往并不关注,实际效果很不理想。^[23]交易是相互的,市场主体在提高他人交易成本的同时,也在提高自己交易的成本,减少自己的交易机会,因为交易对方会取消交易或提高要价以弥补被提高的成本。在没有特殊意义的情况下,市场主体一般不愿意以“提高他人交易成本”的方式实施声誉制裁。故此,仅仅从客观效果出发将市场主体的失信惩戒理解为“提高失信主体交易成本”,无法解释市场主体参与失信惩戒的动力,进而难以作出针对性设计以促成联合惩戒机制。

这样一来,失信惩戒只能沦为行政机关的独角戏,因客观上执法资源不够或主观上执法不严导致的有法不依、执法不公又会再次出现。行政机关除了背负更多、更重的执法负担,并不能通过失信惩戒促成讲诚信的交易环境。对此,早有学者指出:“单靠政府制定法律对严重不讲信用构成犯罪的予以打击,已被多年的实践证明并不十分有效,在中国的民法、合同法和反不正当竞争法中都有诚实守信的法律原则,刑法中更有对诈骗等不讲信用的犯罪科以重刑的规定,但都不足以消灭不讲信用的违法和犯罪活动。”^[24]

(四) 不利于社会信用制度的体系化

其一,不利于市场主体实施的“市场性惩戒”与行政机关实施的失信惩戒的体系化。默认甚至鼓励市场主体基于信誉逻辑实施声誉制裁,并将之作为行政黑名单与通报批评等声誉罚的有机组成部分,在实质上破坏了行政处罚权的国家垄断性。而为了在名义上维护法律制裁的国家垄断性,当下信用立法又无法承认市场主体的失信惩戒权,只能将之归为“提高交易成本之举”。这就人为地割裂了市场主体与行政机关的失信惩戒,也割裂了行政机关发动的声誉罚。

其二,不利于各层次、各类型失信惩戒手段的体系化。按照行政处罚的思路,加强监管措施、限制给付类措施与不面向社会公开的黑名单不属于行政处罚,而其他的失信惩戒措施都属于行政处罚。^[25]这样一来,一些失信惩戒因不属于行政处罚而无法得到行政处罚体系的规范与救济。而且,失信惩戒本身也被人为地分割为需要以行政处罚体系来规范、救济的行为,与不需要也不能以行政处罚体系来规范、救济的行为。失信惩戒的体系化与一体法治化无从实现。

其三,割裂失信惩戒与信用信息,割裂失信信息与其他信用信息。将失信惩戒视为行政处罚,意味着信用信息是失信的事实证据亦即失信责任的构成要件事实,这就是所谓信用信息的“利用”。^[26]按

[21] 林均跃:《论公共和市场两种不同类型的失信惩戒机制及其互补关系》,载《征信》2022年第1期,第25页。

[22] 参见《上海市社会信用条例》第26条、《南京市社会信用条例》第48条、《江苏省社会信用条例》第49条。

[23] 参见梁馨:《餐馆分级形同虚设》,载《美食》2012年第7期,第49页。

[24] 钟楚男主编:《个人信用征信制度》,中国金融出版社2002年版,第12页。类似观点参见王征宇:《美国的个人征信局及其服务》,中国方正出版社2003年版,第3—6页。

[25] 谭冰霖:《处罚法定视野下失信惩戒的规范进路》,载《法学》2022年第2期,第36—38页。

[26] 参见《南京市社会信用条例》第四章与《湖北省社会信用信息管理条例》第四章。

照当下的信用立法,信用信息不仅包括违法、违约类失信信息,还包括基础信息、良好信息与守法、守约信息。它们都应当被“利用”以决定信用奖励还是惩戒。可是,按照当下仅以违法、违约为失信行为要件事实的逻辑,失信惩戒中对守法、守约信息的利用体现在哪里呢?而且,将违法、违约信息视为失信信息,还破坏了信用信息的整体性,泯灭了其他信用信息的价值,使得失信惩戒无法回应实践的多样性。人是社会关系的总和,必然呈现出多面向、动态化的信用形象,其信用必须结合具体领域、综合所有相关信息进行场景式判断。比如,一个交通违章的小店主未必是惯于以次充好的奸商,一个沉默寡言不善沟通的司机也可能是面冷心热的好汉。而要件涵摄式“有失信信息就惩戒”的行政处罚模式,只会造成实践中广泛的误击与冲突,频生诸如见义勇为者遭受交通处罚后被联合失信惩戒的怪象。究其根本,正是因为行政处罚式失信惩戒背离了“信用数据—信用评价”模式,割裂了信用数据与失信惩戒的关联,丧失了基于大数据与人工智能而获得更精确信用评价的可能性。

其四,割裂失信惩戒与信用修复。信用修复为失信惩戒提供了必要的退出程序,^[27]两者理应一体化建构。信用修复往往采取以良好信息中和不好信息,以废除不好信息消除其对信誉的伤害的方式,来证明失信者还是可以挽救的好同志。^[28]一旦信用得以修复,失信惩戒措施就应该减少或降低烈度,但这一逻辑无法协调处罚化的失信惩戒。因为后者旨在追究失信责任并由此填补失信者对信用的危害,而信用修复既不能降低失信者已经造成的社会危害,也不基于失信者对社会的重大贡献。进言之,信用修复措施并不能满足减罚(刑)所要求的功过相抵。^[29]因此,将失信惩戒视为行政处罚,无法在失信惩戒与信用修复之间建立有机统一的联系。

上述困境源于现行信用立法与学理将违法与违约视为失信,将失信上升为法律上的失范,进而将失信惩戒视为制裁、矫正失范行为的法律制裁。其在实质合法性上必然要经由道德法律化路径,在行为类型上必然与行政处罚混同,进而以过惩相当为原则,以公共机关垄断惩戒权。然而,道德法律化的严格条件又决定了这种法治化进阶的诸多困境:难以解释行政机关进入私法领域对违约行为施加强制性制裁的问题;难以解释其他行政机关与市场主体参与联合惩戒的合法性、动力问题;难以解决由此带来的过惩失当、救济无力问题;难以解决行政机关不得不垄断惩戒权进而不得不独自承担惩戒重负,以致难以充分治理失信的难题。不仅如此,行政处罚化还导致失信惩戒被割裂、信用信息被割裂、失信惩戒与信用修复之间的逻辑冲突等多方面问题,使社会信用制度难以围绕着信用而体系化。其认识论与方法论的错误根源在于从道德与规范的角度来界定失信,混同“信用”与“诚实守信”,从而导致失信惩戒制度建构上的根本性方向偏差。

二、服务于资源配置风险之管控的失信惩戒

要避免上述困境,应当回归失信惩戒的本源,从管控、规避失信风险而不是制裁失信的角度来理解、建构失信惩戒。这一进阶的实质是通过信用机制的事前预防机能去管控将来的交易风险、避免失信行为的产生,而不是通过事后的制裁去报复、威慑失信主体使其今后不敢再犯。这一逻辑可以归为失信治理的预防论范畴,与属于制裁论范畴的行政处罚化思路迥然有异。对此,有必要先回溯经济领域的信用、失信与失信惩戒,阐释风险管控面向的失信惩戒的失信治理逻辑,再探讨行政机关为何可以运用失信惩戒。

[27] 参见解志勇、王晓淑:《正当程序视阈下信用修复机制研究》,载《中国海商法研究》2021年第3期,第4页。

[28] 参见戴昕:《信用如何修复》,载《中国法律评论》2021年第1期,第134—137页。

[29] 以刑罚为例,减刑(责)以社会危害性降低和社会危害后果的降低为圭臬,而有重大立功表现是减刑的绝对实质条件,即对社会有重大立功必然减刑。参见陈兴良:《本体刑法学》,商务印书馆2001年版,第850页。

（一）经济领域的信用、失信与失信惩戒

在经济学领域,信用并没有被注入“诚实守信”这一规范内涵。有学者考证指出,“信用”曾一度只适用于信贷领域,意指可以先获得贷款之后再还的能力。^[30]认为借方有信用,即意味着贷方信任该人,“授信”予其没有风险;认为借方会“失信”,即意味着贷方认为借方会使自己陷入不能获得偿付的窘境。为了避免这种风险,贷方在交易前必然要评估借方会否失信,信用由此成为评判交易可信用或安全度的标准:“在信贷领域,……信用评分成为唯一重要的事情。”^[31]信用评分需要计算的是交易对方给特定商业交易带来的特定风险等级。^[32]随着经济与信用评估技术的发展,信用扩展至保险、信贷、合同、消费、儿童收养、行政许可与政府福利等诸多领域,^[33]本质依然是基于信任配置资源。正如有论者指出:“个人职业信用也是经济信用的一部分,如果不对金融从业者、中介机构从业者等重点职业人群以及欺诈等恶劣职务违法违约行为进行权威的、经过反复检验的披露,那么政府、企业与公众均没有可靠标准和有效信号来防范市场交易风险。”^[34]

这就意味着,信用在本源上不是对过往行为的道德定性,而是对将来交易的风险评级,服务于将来交易的安全需要。信用的道德规范维度源于、依附于信用的资源配置维度,两者是上层建筑与经济基础的关系。换言之,信用对于资源配置至关重要,讲信用才具有了道德规范层面的意义。正因为如此,无论是2014年国务院颁布的《纲要》,还是2022年3月29日中共中央办公厅、国务院办公厅印发的《关于推进社会信用体系建设高质量发展促进形成新发展格局的意见》,都是从交易安全维度来理解信用的功能与意义的。前者提出,建设社会信用体系,“是整顿和规范市场经济秩序、改善市场信用环境、降低交易成本、防范经济风险的重要举措”;后者提出,“进一步发挥信用对提高资源配置效率、降低制度性交易成本、防范化解风险的重要作用”。

为了避免陷入失信风险,资源出让方必然会根据信用风险等级采取如下被称为失信惩戒的措施:或是对交易对方进行更为严格、频繁的审查以排除风险,或是要求其提供担保以确保交易安全,或是提高交易对价以对冲风险,或是将之拒之门外不予交易以规避风险。尽管这些措施给交易对方带来了不同程度的不便与麻烦,但并不构成法律上的限制与侵害。因为,这些措施背后是交易双方之间的霍菲尔德式“特权(自由)一无权利”关系:^[35]我有不与你交易的特权(自由),你没有要求我与你交易的权利;我没有强制你按照我提出的条件与我交易的权利,你有拒绝根据我提出的条件与我交易的特权(自由)。

之所以不一律采取不予交易之举而是视情况采取不同的风险管控措施,是因为出让方需要平衡交易安全与交易机会,不能因噎废食地为了绝对的安全而排除一切蕴含着风险的交易机会。为此,信用评分必须结合交易的具体场景,尽可能地收集更多信息作综合判断并决定风险管控形式,而不能仅仅因为交易对手方此前有违法、违约行为就认定失信。我国地方信用立法规定“基础信息、良好信息、失信信息和其他信息”,不只规定“失信信息”,^[36]正是因为这些信息都有助于科学评判交易风险。在传统信用评估中,不独违法、违约信息,企业的品德(character)、资本(capital)、能力(capacity)、抵押(collateral)方面的信息,能够对企业偿付能力产生影响的社会经济发展趋势与特定发展和变动情况(condition)信

[30] 林均跃:《社会信用体系原理》,中国方正出版社2003年版,第13页。

[31] F. Pasquale, Restoring Transparency to Automated Authority, 9 Journal on Telecommunications High Technology Law 235, 248 (2011).

[32] See Ian O'Neill, Disparate Impact, Federal/State Tension, and the Use of Credit Scores by Insurance Companies, 19 (2) Loyola Consumer Law Review 151, 152-153 (2006).

[33] 参见《公平信用报告法》第1681条b款。

[34] See Scott R. Peppet, Unraveling Privacy: The Personal Prospectus and the Threat of a Full-Disclosure Future, 105 Northwestern University Law Review 1153, 1168 (2011).

[35] See W.N. Hohfeld, Fundamental Legal Conceptions as Applied in Judicial Reasoning 39-50 (Walter Wheeler Cook ed., Yale University Press 1919).

[36] 参见《南京市社会信用条例》第20条第3款。

息,个人身份、信贷与公共事业缴费信息、公共记录即遵纪守法信息、就业信息与消费信息,都属信用信息。^[37]就此而言,基于信用评分而安排交易策略,本质上就是方兴未艾的“基于数据的治理”。基于这一逻辑,在大数据时代,在“所有数据都是信用数据”的理念指导与强大的数据收集、处理技术的加持下,信用信息的范围更是无远弗届,人们的购物地点、购物行为、在线社交媒体网络以及与信誉没有直观关联的其他因素都被囊括其中。^[38]面对如此复杂多样的信息,信用评分更加不可能采用“有违法、违约就算失信”的要件构成模式,而是必须依赖于越来越复杂的信用算法。

综上,经济领域中的失信惩戒是交易主体为了交易安全,基于信用信息,评估交易对方的信用后作出的风险管控举措。当然,在管控、规避交易风险的同时,失信惩戒在客观上具有不容否定的惩戒效果,即它让失信者失去了交易机会或是要付出更多的成本才能获得交易机会。而且,这种机制越是能预防失信,对失信行为的惩戒效果就越充分、越强烈。基于预防失信、规避风险的需求,每个市场主体必然会无需动员地参与其中,形成一个无组织的社会性制裁。^[39]也就是说,不是因为失信主体的行为有违公德、伤害了多少市场主体,所以市场主体出于维护正义或是报复的动机一起惩罚他,而是因为失信主体对他们来说都有失信可能,他们出于管控、规避风险的目的,才积极参与到这个预防机制中,对与其“无冤无仇”的失信主体实施惩戒。失信者正是在被联合抵制、处处得不到交易机会的意义上“一处失信,步步受限”。就此而言,只有从管控风险这一动机出发,才能理解为何市场主体会主动参与联合失信惩戒,哪怕这样提高了自己的交易成本,减少了自己的交易机会;才能理解失信惩戒为何只有成为“社会”信用惩戒才真正具有惩戒效果,才成为失信治理机制。^[40]进一步说,正是通过社会联防机制,以“不予交易机会”“严格审查”或“对失信者提高对价、要求其提供担保否则不与之交易”为基本形式的失信惩戒,才有促成诚信交易的威力,才成为现代市场机制不可或缺的装置。

因为社会信用机制能动员市场主体形成社会化制裁,起到激励市场主体更佳守法、优化合规水平的作用,所以行政机关越来越喜欢通过公布违法事实或黑名单实施监管。这不仅有效,而且还因为“更少强制、更多声誉”,^[41]被视为“行政机构寻求推动有意义变革的阻力最小的途径”。^[42]因为,“信息披露符合自由市场和自治原则,与更传统的监管相比,似乎是一种低成本、简单、有效的柔性干预”,^[43]由此,“国家或政府集中供给的直接规制完全可以限缩,让位于各领域中自发生成的声誉机制,由此系统性地降低管理成本,提高管理效率”。^[44]总之,趋利避害心理会驱动市场主体自觉参与、执行失信惩戒,失信惩戒机制由此成为一个只需国家投入少量资源加以润滑就可以运行无碍的系统,可以高度自足地生成经济领域信用环境。

为了使基于风险管控心理的市场信用治理机制更好的运转,同时也为了提升市场主体管控交易风险的水平,经济领域信用体系的核心应当是信用信息的公开机制,而不是国家的失信惩戒机制。正如张维迎等经济学家所指出的:信用问题的核心是信息问题,只要有足够的信息,就很少会发生不守信用的问题;若是信息不充分,欺诈行为不能被观测到的话,信用就会出问题。^[45]因此,发达国家和地区通常并不特别详

[37] 林均跃:《社会信用体系原理》,中国方正出版社2003年版,第73—80页。

[38] Mikella Hurley & Julius Adebayo, Credit Scoring in the Era of Big Data, 18 Yale Journal of Law and Technology 148, 151 (2016).

[39] 关于社会性制裁,参见[日]佐伯仁志:《制裁论》,丁胜明译,北京大学出版社2018年版,第6页。

[40] 林均跃:《社会信用体系原理》,中国方正出版社2003年版,第53页。

[41] Lior Jacob Strahilevitz, Less Regulation, More Reputation, in The Reputation Society: How Online Opinions Are Reshaping the Offline World 63 (Hassan Masum & Mark Tovey eds., The MIT Press 2012).

[42] See William M. Sage, Regulating through Information: Disclosure Laws and American Health Care, 99 Colum. L. Rev. 1701, 1772 (1999).

[43] Omri Ben-Shahar & Carl E. Schneider, The Failure of Mandated Disclosure, 159 University of Pennsylvania Law Review 103, 134—135 (2011).

[44] 戴昕:《理解社会信用体系建设的整体视角:法治分散、德治集中与规制强化》,载《中外法学》2019年第6期,第1474页。

[45] 参见张维迎:《信息与信用》,载《市场营销导刊》2005年第2期,第9页。

尽可能地规定信用联合激励和联合惩戒的具体法律规则,而更多地留给市场经济主体根据交易的情况去实施信用激励和信用惩戒。^[46]为此,我国应当建构一个在合法合规的前提下充分开放各种信用信息的社会信用机制。一切资源配置主体都可以根据各自风险管控需求,各取所需信用信息,自行判断失信风险,进而决定是否实施失信惩戒,并基于相同需求形成特定交易领域的自发性社会联合惩戒。而行政机关则不需要也不应该基于国家垄断性暴力逻辑,亲自下场对经济交易领域中的失信行为实施行政处罚式失信惩戒。

(二) 行政机关的失信惩戒

行政机关不应强制性制裁经济交易领域中的失信,但可否实施作为管控资源配置风险之举的失信惩戒呢?答案是肯定的。

1. 符合行政机关管控资源配置风险的需要

一方面,行政机关要利用财政资金与国有土地等公共资源,从市场主体那里换取各种各样有形、无形的商品,因而也需要通过信用评价机制来判断风险,进而决定该如何配置公共资源才能确保避免市场主体违约;也可以将失信界定为市场主体不足为信,向其配置公共资源存在风险。因此,行政机关也可以像市场主体那样,出于管控风险的动机,实施诸如审慎调查或加大检查频次,要求提供担保或提高交易对价,排除特定市场主体缔约资格等失信惩戒措施,并和其他有同样需求的交易主体形成自发的联合惩戒。这和市场主体的失信惩戒一样具有充足合法性。当然,因为涉及公民对于公共资源的公平分享权,行政机关不能像市场主体那样,以契约自由为名不受限制地实施失信惩戒。^[47]

另一方面,行政机关也可以出于管控风险的动机在行政监管领域实施失信惩戒。正如美国制度经济学家康芒斯指出的,所有的,无论是家庭的、商业的或政治的各种经济和社会关系的互动都是交易,而且,每一个交易都有期望的经济维度与心理维度。^[48]在这个意义上,行政执法行为也是要追求效益与安全性的资源配置行为。为了避免执法资源的错配,行政机关在执法领域也可以根据相对人的各种相关信息实施失信惩戒,亦即根据信用评估决定监管策略,并形成这个意义上的联合惩戒。当然,行政机关执法的最终目的是保障社会秩序,服务于社会大众,故不能仅仅出于避免“出事”的动机,粗暴地实施一刀切式失信惩戒。

归根结底,为了平衡社会大众的自主交易需要与安全交易需要,行政机关需要采取不同的失信惩戒来管控风险:适用提高检查频次、适用审慎检查程序等强化监管措施而不是适用简易审查或是告知承诺制程序等失信惩戒,是为了避免、预防相对人因为监管资源配置不到位、脱离监管视线而违法。其之所以“轻微”,是因为相对人违法的可能性小或是危害小,采取不禁绝相对人交易自由的措施即足以应对。采取从业限制等“严重”的失信惩戒,事先断绝信用不足的相对人进入特定市场的自由,是因为他们违法可能性很大或危害很大,不如此不足以预防风险。采取通报批评、公布违法事实等烈度居中的失信惩戒,则是通过广而告之的方式,提醒社会大众注意并预防交易风险。它们在更大范围内提高了相对人的交易成本,却没有禁绝他们的交易自由,是因为他们违法可能性较高或是危害较大,但又没到以断绝交易自由的方式来预防的程度。综上,严格审查、提高检查频次等措施,与通报批评、公布违法事实及从业禁止等失信惩戒措施,共同构成一个体系化、有层次的执法资源配置风险管控体系。^[49]就此而言,与旨在通过事后制裁来治理失信行为的行政处罚化思路相比,从风险管控视角出发,我国当下信用立法所建构的“公法化”失信惩戒体系,可以得到更为体系化、层次化的解释与建构,并由此具备一体法治化的契机。

正是出于管控风险的需要,如今被称之为失信惩戒的措施早已被广泛运用于行政管理领域。美国行政机关一直像市场主体那样获取并根据信用报告,决定是否给予行政相对人许可、执照或福利,是否

[46] 王伟等:《企业信息公示与信用监管机制比较研究:域外经验与中国实践》,法律出版社2020年版,第21页。

[47] 参见陈国栋:《缔约资格排除类信用惩戒的法治化路径》,载《现代法学》2021年第1期,第112页。

[48] [美]康芒斯:《资本主义的法律基础》,戴昕等译,华夏出版社2009年版,第5、87、77页。

[49] 更为详细的阐释,参见陈国栋:《论违法行为的信用惩戒》,载《法学评论》2021年第6期,第44—49页。

允许其进入特定领域或市场,且这种决定与市场主体的决定被《公平信用报告法》等同视之。^[50]我国虽然直到2014年才出台《纲要》,但基于信用评估的行政管理也早已有之。被包括我国刑法、公司法在内的诸多公法、私法所规定,且作为我国当下失信惩戒之主要类型的从业禁止就是其中典型。而加大检查频次、适用严格审查程序等失信惩戒类型在实践中更是俯拾皆是,其中一个重要原因就是这样的避免常规化配置执法资源所致的政府目标难以实现的风险。^[51]《国务院办公厅关于加快推进社会信用体系建设构建以信用为基础的新型监管机制的指导意见》(国办发〔2019〕35号)所提出的“大力推进信用分级分类监管。在充分掌握信用信息、综合研判信用状况的基础上,以公共信用综合评价结果、行业信用评价结果等为依据,对监管对象进行分级分类,根据信用等级高低采取差异化的监管措施”,更是体现了根据信用评级分配执法资源、规避执法资源配置风险的思路。

2. 符合行政机关与市场主体联合共治的要求

给社会提供更安全的交易环境是行政机关的基本职能,也是行政机关实施失信惩戒的初衷,更是行政机关与市场主体联合共治的基础。但如果行政机关只依赖单方面的暴力机制,就难免因独木难支而无法充分实现这一目标。为此,行政机关必须尽可能与市场主体联合共治,以避免单兵作战的困境。以管控交易风险为目的实施失信惩戒,正好能营造政府与市场主体同频共振治理失信的局面,因为规避风险是驱动市场主体与行政机关切实合作的公约数。

一方面,市场主体为了更好地预防失信,会无需组织地参与到联合惩戒中来,与行政机关一起形成“面对点”的失信治理机制。通过信息披露的行政监管之所以能实现更有效的监管,正是因为行政机关发布信用信息的目的不是为了惩戒,而是为了提示交易风险。这就激发了市场主体实施失信惩戒的动力,形成了行政机关与市场主体在失信惩戒上的有机联合,避免了行政机关独自发动“点对点”式法律制裁的窘境。如果行政机关按照惩罚的逻辑披露信息或实施其他失信惩戒形式,那其与市场主体的失信惩戒就必然被分割为法律制裁与所谓的“提高交易成本的举措”。两者不但难以织就社会性联防之网,反而会因逻辑不一致、动力不一致而网开多面。

另一方面,市场主体为了更好地规避风险,还会主动提供各种信用信息以支撑、扩张信用机制作用的领域、场景与有效度。^[52]也就是说,市场主体不仅通过失信惩戒规训参与资源配置的主体,使其不敢、不能失信,还会提供各种信用信息便于包括行政机关在内的资源配置主体实施惩戒。网络购物中的差评与好评即是这一逻辑的典型样态。归根结底,失信惩戒以信息的收集、评估为基础,信息越丰富,评估就越可能准确,风险就越能得到理性的预判与管控。故此,失信惩戒只有从管控风险的角度来理解、建构,才能激发市场主体提供信用信息的积极性,为广泛的联合共治建立坚实的信息基础。

3. 符合大数据时代社会治理机制的变革趋势与创新需求

众所周知,随着社会风险的日益增多与越发复杂,传统社会治理机制越来越捉襟见肘。为此,国家不得不寻求创新社会治理方式,以便以有限的资源实现更有效的社会治理。就此而言,广泛适用以管控资源配置风险、优化资源配置效益的失信惩戒,是国家治理的客观需要。然而,失信惩戒在很长时间内没有得到广泛适用。预防论的刑法并非刑法主流,以预防违法为主要目的的诸多行政强制措施只能依附于、混同于行政处罚进而被严格限制,即为鲜明例证。^[53]究其根本,因为这会否定或限制、剥夺公民的

[50] 参见美国《公平信用报告法》(FCAR)第1681条b款。

[51] 黄镛:《为什么选择性执法?制度动因及其规制》,载《中外法学》2021年第3期,第787页。

[52] See Larry Catá Backer, Next Generation Law: Data-Driven Governance and Accountability-Based Regulatory Systems in the West, and Social Credit Regimes in China, 28 Southern California Interdisciplinary Law Journal 123, 138 (2018).

[53] 参见何荣功:《预防刑法的扩张及其限度》,载《法学研究》2017年第4期,第148页;李洪雷:《论我国行政处罚制度的完善——兼评〈中华人民共和国行政处罚法(修订草案)〉》,载《法商研究》2021年第6期,第5—6页。

公法权利,行政机关的失信惩戒只有在审慎精准时才有合法性,不可能像私法信用惩戒那样为私法自治原则所庇护,为市场经济的风险自担、盈亏自负机制所动态调整。比如,如何确定是否可以对相对人实施从业禁止、缔约资格排除等失信惩戒?如何确定其实施期限是1年、3年、5年还是终身?如何保证它们不是简单粗暴的,不是“不确定的、难以捉摸的”?^[54]在信息收集、处理能力不够强大,不能确保信用评估科学、可靠的情况下,行政机关自然应当慎用失信惩戒。

但大数据与人工智能有效地打破了上述桎梏,支撑失信惩戒成为政府有效应对繁多复杂的社会风险的治理手段。^[55]大数据的核心就在于预测,数据越多,预测就越准确,^[56]其与基于大量数据而预测、应对失信风险的失信惩戒具有天然的亲和性。在大数据与人工智能的辅助下,行政机关可以在社会问题还未达到明显、重大、突出之时就迅速预测出来,并根据相对人违法风险的等级与危害程度,有针对性地配置执法资源消除、隔离违法风险,而不是在违法之后再制裁。这种“按风险大小分配资源”的“基于风险的规制”(risk-based regulation),可以让行政机关实施“有的放矢”“松紧有度”的干预,在“最大化监管收益”的同时“最小化被监管方的负担”,^[57]精准实现执法资源配置的优化与治理的优化。而且,在大数据与人工智能算法的优化下,存在巨大裁量空间的失信惩戒也会在合理性、可靠性方面得到有效规范与提高。正是基于这两方面的效用,2019年1月发布的《国务院办公厅关于全面推行行政执法公示制度执法全过程记录制度重大执法决定法制审核制度的指导意见》第17条提出:“一是要推动行政执法裁量的决策辅助系统,规范行政自由裁量权,确保执法尺度统一;二是要开发风险预测系统,提前预警、检测、研判,及时履行政府职能、提高行政精准性。”

随着政府像平台公司一样平台化、数字化,失信惩戒必然迅速扩张于各个治理领域,取代过往以“命令—制裁”为核心的社会治理机制。“算法已经开始并且将不可避免地以不可想象、史无前例的方式驱动包括公共管理的组织形式在内的社会组织形式发生激烈变革”,已成为政治学家的共识。^[58]拉里·卡塔·巴克(Larry Catá Backer)更是明确指出,随着新的数字技术日益弥漫、支配经济生产乃至相应的上层建筑,一种全新的数据驱动的治理机制会取代过往的法律机制:“在这个新世界中,旧式的法律、国家和秩序可能会有一席之地,但它必然会让路给以测量、评估和激励为基本意识形态的制度体系。”^[59]在算法驱动下,法律的介入路径将由“行为后”转向“行为前”,即由聚焦“如何修复和恢复社会”转向“如何规制和塑造社会”,由“裁断行为后果”前移为“塑造行为逻辑”,也就是前移到对社会关系“架构”的建设上来。^[60]就此而言,大数据与人工智能不仅为失信惩戒提供了坚实的技术支撑,也在推动它成为主流的“基于数据”的治理方式。而将之理解为行政处罚,只会使之因背离大数据时代社会治理机制的变革与发展趋势而边缘化。

失信惩戒还可以作为大数据与人工智能时代的治理术与规训机制。^[61]正如传统的“命令—制裁”式社会治理机制不仅依靠国家垄断的暴力,更依靠公民的守法观念,因而国家要以各种方式提升公民的守法意识一样,以信用治理为核心的数字化治理机制也需要失信惩戒作为相应的规训机制提升公民的守信意识。基于越是能规训自己符合信用体制的要求就越能获得资源的信用法则,公民不得不遵纪守法、诚实守信,守信

[54] 何荣功:《预防刑法的扩张及其限度》,载《法学研究》2017年第4期,第145页。

[55] 参见曾光辉编著:《信用大数据时代:赋能社会治理、金融科技与数字经济》,中国工人出版社2021年版,第29页。

[56] [英]维克托·迈尔-舍恩伯格、肯尼思·库克耶:《大数据时代:生活、工作与思维的大变革》,盛杨燕、周涛等译,浙江人民出版社2013年版,第16页。

[57] Alex Griffiths, *The Practical Challenges of Implementing Algorithmic Regulation for Public Services*, in *Algorithmic Regulation* 144, 152 (Karen Yeung & Martin Lodge eds., Oxford University Press 2019).

[58] Karen Yeung & Martin Lodge, *The Practical Challenges of Implementing Algorithmic Regulation for Public Services*, in *Algorithmic Regulation*, 1, 1 (Karen Yeung and Martin Lodge ed, Oxford University Press 2019).

[59] Larry Catá Backer, *Next Generation Law: Data-Driven Governance and Accountability-Based Regulatory Systems in the West, and Social Credit Regimes in China*, 28 *Southern California Interdisciplinary Law Journal* 123, 172 (2018).

[60] 齐延平:《数字化社会的法律调控》,载《中国法学》2022年第1期,第88页。

[61] 参见虞青松:《算法行政:社会信用体系治理范式及其法治化》,载《法学论坛》2020年第2期,第36—37页。

由此得以内化。而信用激励与信用惩戒正好既一体两面、无法分割,又有机地此消彼长。在交易需求不变的情况下,因失信惩戒而取消的交易必然要流向守信者,守信者必然因获得更多交易机会而更加守信。由此,基于信用机制所内含的激励与惩戒效应,通过“定义、量化和计算可信度和诚实度”的方法,国家运用失信、守信的二元编码,将符合信用逻辑、旨在追求利益最大化的公民塑造成新的“文明”(“有信用意识”)人。^[62]

三、失信惩戒法治化的建构

大数据与人工智能算法补足了失信惩戒精准性欠缺的短板,提升了它的合法性,推动其成为主流的社会治理方式,也给它的法治化带来了难题。因为,“算法在很多方面危及包括法治与基本权利在内的多数基石性法律概念,动摇了现存法律体系”。^[63]换言之,现存的法律框架难以胜任失信惩戒法治化的重任,需要为之设计全新的法治化架构。限于主旨与篇幅,本文要全面展开这样一个命题难免挂一漏万。但基于失信惩戒的数据治理手段的本质、风险管控属性及其社会性,还是可以从信用算法而非失信惩戒行为入手,以信用算法为中心建构法治化失信惩戒的大致框架。

(一) 以信用算法作为失信惩戒法治化的中心

失信惩戒旨在管控将来的交易风险,因此信用算法是典型的预测算法。在大数据时代,预测算法是通过分析大数据来揭示(人、环境或活动)各种特征与期望的或不期望的结果之间的相关性……具有最佳匹配性的相关性设定将被建模,这种模型能够在给出相关联的事实后确定将来这一事件的大概走势及发生的可能性。^[64]就此而言,信用算法就是按照一定的相关性设定,对特定信用信息进行处理,从而预测相对人的失信风险及其等级以及相应应对方式的方法。以信用算法而非惩戒行为作为失信惩戒法治化的中心,主要出于如下三方面原因。

其一,以行政行为合法性为中心的依法行政机制无法有效规范失信惩戒行为。以行政处罚为例,传统依法行政机制主要审视其是否具备法律授权、是否遵守正当程序与是否基于已经查实的违法行为事实。失信惩戒则无法被纳入这一框架。以缔约资格排除为例,信用算法根据市场主体的各种信息,预测其违约的可能性很高或是不能承受其违约风险,就会将其列入不得与之缔约的黑名单。但是,直到参与招投标被拒之前,该市场主体都不会知道其已经被列入黑名单,更不会知道原因与过程。一是行政机关不是在强制性地剥夺该市场主体的自由与财产,而是出于管控风险、保全自己的考虑而不向其配置资源,所以不需要通过法律授权与正当程序获得正当性;二是行政机关可以通过征信机制、政府数据开放机制或其他公私合作渠道获得信用信息,不需要通过正当程序满足信息收集的正当性;三是风险预判与算法运作属于内部行政行为,无法受到正当程序规制。^[65]总之,传统行政行为合法性机制对其无从下手、无法评判。

其二,失信惩戒的合法性在于信用算法的精准性。失信惩戒的基础是交易风险评估,只要信用算法评定存在交易风险,失信惩戒就有合法性。不规范失信惩戒赖以发生的信用“画像”算法,^[66]就可能否定惩戒措施的合法性。在算法就是法律的时代,^[67]就像恶法不断产生恶行一样,不科学、不公平的算法

[62] Nicholas Loubere & Stefan Brehm, *The Global Age of the Algorithm: Social Credit and the Financialisation of Governance in China*, p. 144, https://www.researchgate.net/publication/332232338_The_Global_Age_of_the_Algorithm_Social_Credit_and_the_Financialisation_of_Governance_in_China (accessed Jan. 8, 2023).

[63] Karen Yeung & Martin Lodge, *The Practical Challenges of Implementing Algorithmic Regulation for Public Services*, in *Algorithmic Regulation* 1, 2 (Karen Yeung & Martin Lodge eds., Oxford University Press 2019).

[64] Robert Brauneis & Ellen P. Goodman, *Algorithmic Information Disclosure by regulators*, 20 *Yale J.L. & Tech.* 103, 113 (2018).

[65] 张凌寒:《算法自动化决策与行政正当程序制度的冲突与调和》,载《东方法学》2020年第6期,第7页。

[66] [美] 卢克·多梅尔:《算法时代:新经济的新引擎》,胡小锐、钟毅译,中信出版集团2016年版,第12—27页。

[67] [美] 劳伦斯·莱斯格:《代码2.0:网络空间中的法律》,李旭、沈伟伟译,清华大学出版社2009年版,第6页。

会源源不断地产生错误的“惩戒”，这将导致不合理失信惩戒的体制化、系统化、持续化与普遍化。^[68]比如，美国传统信贷风险评级算法在帮助银行管控信贷风险的同时排斥了一些群体获得信贷的机会，而这样一个群体，经由另一种算法却被认为是可以获得银行贷款的。^[69]因此，必须超越后果审查式行为规制，将规制重心前置于算法运作环节，对失信惩戒实行釜底抽薪式规制。

其三，失信惩戒的社会性也决定了要以信用算法为法治化中心。失信惩戒是在风险管控动机下驱动的，其要害不在于最后的惩戒，而在于信用风险的评级环节。一旦负面的行政信用评价出台，同一领域的资源配置主体但凡知晓该负面评级，就会出于风险管控需要实施自发性联合惩戒，形成“一处受限，处处受限”的场景。在这种情况下，法律很难逐个规制惩戒主体。故此，为避免不精准的信用评级生成后覆水难收，就必须以信用算法为失信惩戒法治化之中心。

当然，规制信用信息也是规制信用算法的重要路径。传统失信惩戒的规制路径就是审查信用信息的真实、准确与否，如《美国公平信用报告法》的主旨之一就是确保消费者免受不准确信息的影响。^[70]西方学者近年来以删除权、被遗忘权、不联网权来解决大数据时代错误信息、隐私信息被用于社会信用评价的问题，^[71]也可以置于这一脉络来理解。但是，由于信用信息服务于信用算法的运作，因此，对信用信息的规制可以纳入信用算法的规制。对此，下文将予以展开。

（二）以风险管控原则规范信用算法

既然失信惩戒的合法性在于风险管控，那么它就必须以此为限度与边界。这就意味着，风险管控不仅是失信惩戒的目的，也是其必须自始至终应予遵守的根本原则。作为失信惩戒的基础，信用算法自然也不能逃离这一原则的规范。对此，可以从如下几个层层推进的阶段加以把握：

首先，风险管控原则意味着信用算法必须去道德化或意识形态化，不以维护和实现特定的道德或价值偏好为目标。基于我国宪法第24条规定的精神文明条款，国家机关当然要维护、宣扬、落实社会主义精神文明，但当下一些地方信用立法将失信惩戒作为推进社会公德之手段的做法应当戒除。^[72]一方面，道德化意味着制裁。其若主导失信惩戒，就会引发失信惩戒的性质变化及相应的合法性危机。另一方面，道德化和交易风险的风险管控需求不协调。如前所述，行政机关应该作为资源配置主体实施失信惩戒，这样才有可能因同质于市场主体的风险管控需求而形成公私联合的社会信用惩戒。道德化则有碍于此，因为其会提高交易成本，阻碍交易的发生，所以道德化的信用算法不仅得不到市场的配合与适用，反而会成为社会信用体系的短板，损害这一体系的稳定性与权威性。基于这一逻辑，社会信用体系无需区分守信、失信信息，只需将各种依法可供查询的信息归集到信用数据库即可。

其次，风险管控原则意味着信用算法要以包容为原则。有些学者提出失信惩戒要奉行谦抑原则。^[73]不过，从管控风险而不是惩戒失信这一目的出发，将其命名为包容原则或鼓励交易原则更为合理。失信惩戒是为了管控风险而不是单纯地限制交易、规避风险。纯以规避失信为倾向，既会造成实践中业已出现的失信惩戒泛滥、公民动辄得咎的乱象，还会形成过度防御，使得被惩戒者因交易能力越发削弱而陷入反复失信、愈发失信的恶性循环。^[74]这就意味着信用算法必须更具包容性，以最大化保留交易自由、

[68] 参见马长山：《数字时代的法律变革与权利保护》，法律出版社2021年版，第124页。

[69] [加]雷蒙·安德森：《信用评分工具：自动化信用管理的理论与实践》，李志勇译，中国金融出版社2017年版，第22页。

[70] 参见《美国公平信用报告法》（FCAR）第1681条a款。

[71] See Nizan Geslevich Packin & Yafit Lev-Aretz, On Social Credit and the Right to Be Unnetworked, 2 Columbia Business Law Review 339, 420 (2016).

[72] 《苏州上线“文明码”引争议 当地回应：以市民自愿注册为前提》，载央广网，https://china.cnr.cn/yaowen/20200908/t20200908_525244655.html，最后访问日期：2023年1月8日。

[73] 顾敏康、谢勇、王伟、石新中：《我国诚信建设法治化核心命题笔谈》，载《求索》2020年第3期，第16页。

[74] F. Pasquale, Restoring Transparency to Automated Authority, 9 J. on Telecomm. & High Tech. L. 235, 237 (2011).

扩张交易机会为倾向,否则就可能导致系统性的歧视与不公平对待。^[75] 正如有论者指出的,“一个人偿还贷款的相对能力可能取决于多种因素,而一个更细致的、能整合更广泛的数据的模型,至少可以解决包容性不足的问题”。^[76] 行政机关应该基于大数据思维,开发更具包容性、更能平衡风险管控与交易扩张需求的算法,尽量保障市场主体的交易自由以激发市场活力、扩大交易机会。这意味着,信用算法不能停留在一刀切式的有违法、违约行为就认定失信的层次,也不能驻足在首次违法不算失信这种简单处理的层次,更不能满足于传统的违法、违约风险计算模型。

再次,风险管控原则意味着信用算法的类型化。根据算法规制的场景化原则,^[77] 计算失信可能性及其危害程度的信用算法必须符合具体场景、具体领域下风险管控的需求,必须依托失信风险的类型与领域而类型化、具体化。与之相应,信用评级不能跨风险区域而适用,不得具有跨区域效力。这是因为行政机关所面临的风险虽然很多,但不同机关所面临风险不一样,同一机关的不同交易类型所面临的风险也不一样。比如,行政机关无论是配置执法资源还是配置公共资源都需要管控风险,但两者也有区别。因为,执法资源的配置归根结底是为大众的风险管控服务的,而公共资源配置主要服从自身的风险管控需求即可。因此,行政机关应当根据不同的风险规避需要,设计不同的信用算法。正如有学者指出的,“在不同行业 and 不同领域,信用维度各有差异……信用的分级分类由用信主体依据用信场景,运用不同的算法自主确定,是唯一科学的做法。即便是政府,不同部门的信用需求也不相同……”^[78] 也正因为风险的存在决定了失信惩戒的适用,风险的类型及其界定既需要随着时代的发展而发展,也需要维持一定的稳定性,不能由行政机关随意界定并设计、运用信用算法与失信惩戒。为此,进行一定程度的顶层设计,由高层级立法统一规定风险的类型或统一规定风险认定的程序,是必要的。

然后,风险管控原则意味着信用算法所需信息亦即信用信息的归集必须坚持真实性、有限性与关联性。算法以信息为前提,特定的算法需要特定的信息,特定的信息服务于特定的算法及相应的风险管控需要。如网络婚姻介绍即有不同算法——长相、体型、DNA以及基于传统数据挖掘出的经验,所需要的信息可区分为长相、体型、DNA与传统婚介信息。^[79] 这就是失信惩戒和信用信息一体化的根源所在。基于风险管控的需要,行政机关只能收集、利用可用来评判特定风险之有无、大小的信息,而不能过度收集信息,更不能不管信息的真假。而关联性又决定了有限性或最小性。我国个人信息保护法第6条规定:“处理个人信息应当具有明确、合理的目的,并应当与处理目的直接相关,采取对个人权益影响最小的方式。”行政机关一定要认识到,更多的数据并不一定总是更好的数据,质量比数量更重要。^[80] 过度收集不仅于事无补,还会放大个人信息泄露的风险,并为过度复杂的算法提供不透明、不负责的借口。尽管大数据使得“一切数据都是信用数据”,但万物有联的同时万物依然有别,同样的信息对于不同的事物在权重与关联性上别有意蕴。因此,信用信息的范围必须由风险、信用算法的逻辑来限定。此外,基于成本收益分析方法,为了保护更为重要的利益,对于一些敏感的隐私信息,纵使其有利于更为准确的信用评级,行政机关也不得收集并利用。域外有国家即排除了基因信息在医疗保险和就业信用评估中的适用。^[81] 又如,为避免过度影响个人生计、抑制创新与发展的反效果,个人信息若是并未严重至须提示

[75] Alex Griffiths, *The Practical Challenges of Implementing Algorithmic Regulation for Public Services*, in *Algorithmic Regulation* 144, 158 (Karen Yeung & Martin Lodge eds., Oxford University Press 2019).

[76] Mikella Hurlley & Julius Adebayo, *Credit Scoring in the Era of Big Data*, 18 *Yale Journal of Law and Technology* 148, 172 (2016).

[77] 参见丁晓东:《论算法的法律规制》,载《中国社会科学》2020年第12期,第146页。

[78] 罗培新:《社会信用法:原理·规则·案例》,北京大学出版社2018年版,第43页。

[79] [美] 卢克·多梅尔:《算法时代:新经济的新引擎》,胡小锐、钟毅译,中信出版集团2016年版,第63—69页。

[80] N. Cortez, *Regulation by Database*, 89 *University of Colorado Law Review* 1, 11 (2018).

[81] Nizan Geslevich Packin & Yafit Lev-Aretz, *On Social Credit and the Right to Be Unnetworked*, 2 *Columbia Business Law Review* 339, 424 (2016).

他人注意交易、交往风险的地步,是否进入信用档案,值得审慎考量。^[82]

最后,风险管控原则意味着信用算法应当确保惩戒手段与风险的匹配性,即惩戒手段必须是能够管控该风险且是对相对人限制最小的。第一,如果不存在风险,或者风险已有其他手段予以限制而不再有发生的可能,就无需实施失信惩戒。比如,一个人已经年老退休,不会再次从事相关职业岗位并利用该岗位之便利而违法犯罪,也就无需再限制其从业资格;一个人在这个领域有再犯风险,但在其他领域没有,就无需实施其他领域的资格限制。第二,如果交易风险切实存在,那么公共机关就需要根据风险的大小与失信行为的社会危害度分级实施失信惩戒。如果能通过加大检查频次或严格审查等方式管控风险,就不要采取通报批评、面向社会发布黑名单等形式;如果可以通过发布黑名单、通报批评等方式来应对风险,就不要采取从业限制等禁绝交易自由的方式。总之,原则上要尽可能地保障、尊重市场主体的交易意愿与机会。第三,基于风险管控原则,失信惩戒必须进行成本收益分析。如果失信惩戒耗费了更多资源,却只是保障了有限资源的安全,那这种得不偿失的失信惩戒就违背了风险管控原则,是不可取的。毕竟,信息的收集、获取、存储、评估,隐私保障,算法的设计,算力的运用,惩戒的实施,以及惩戒的监督、救济都需要成本。第四,失信惩戒需要在与其他社会治理手段综合比较的情况下运用。进言之,面对治理目标,是否就一定要用失信惩戒而不是其他治理手段?有学者指出,过度适用基于大数据的信用机制,会导致合同与信任等诸多其他社会治理机制丧失机能。^[83]还有学者指出,通过信用机制所实现的“一个没有犯罪的世界值得我们去追求吗?让人不可能犯罪(而不是在事后给予惩戒),我们得到了些什么,又失去些什么呢?这种方法对重要的民主价值,如团结、异议和审慎有什么影响呢?”^[84]为防范这些风险,我们不能不审慎考虑失信惩戒的适用范围与场景。

(三) 以信用算法为中心的失信惩戒法治化的基本框架

“基于数据的治理”需要通过“对数据的治理”来加以规制。既然信用算法与相应失信惩戒是出于预防失信风险的目的而对个人信用信息的大数据化收集、处理与运用,那么基于这一逻辑,可以如下两方面为抓手建构失信惩戒法治化的基本框架。

其一,由公民基于个人信息权益规制信用算法与失信惩戒。这一规制思路,业已在个人信息保护法中有所呈现。虽然该法旨在保障个人信息权益,但也可以用于规制失信惩戒,因为失信惩戒建基于对个人信用信息的收集与处理。该法不仅规定个人信息查询权、更正权、删除权、被告知权,还有自动化决策的解释权以及有条件的可携带权等具体权利以及行政机关运用、处理个人信息所需满足的条件与所需遵循的程序,这些都可用于规范失信惩戒的信息收集、处理、评估与适用环节。根据该法,公民不仅可以依法查询并更正、删除不实或过时信用信息,还可以要求补充相关信用信息;除非法律有例外规定,否则行政机关在作出信用评估与失信惩戒时,必须履行主动告知义务并公开、解释信用算法。在大数据时代,基于正当性逻辑的不同,法律智能系统的说理必须以“解释”替代“论证”。^[85]为此,行政机关需要解释:其基于何种风险而运用信用算法作出信用评级;这种算法的逻辑与科学原理是什么;其运用了公民的哪些个人信息、这些信息与算法的关联及其权重是什么;其是如何表征、赋值、呈现个人信息的;为何其采用的算法是最为优化、包容的算法。这一点尤其重要,因为不同算法的复杂性与原理不一样。^[86]在这个数据大爆炸的时代,仅仅公开失信惩戒的处理结果及相关数据是不够的,因为“公开性和知识

[82] 参见王瑞雪:《论横跨公私法域的个人信用评价制度》,载《暨南学报(哲学社会科学版)》2022年第1期,第75页。

[83] Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, 30 *Journal of Information Technology* 75, 81 (2015).

[84] Evgeny Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism* 194-195 (Public Affairs 2013).

[85] 参见刘东亮:《新一代法律智能系统的逻辑推理和论证说理》,载《中国法学》2022年第3期,第159页以下。

[86] Alex Griffiths, *The Practical Challenges of Implementing Algorithmic Regulation for Public Services*, in *Algorithmic Regulation* 154-158 (Karen Yeung & Martin Lodge eds., Oxford University Press 2019).

之间的相关性,或者信息的可用性和基于信息作出的决定的质量之间的关联性”已经没有充分说服力了,^[87]行政机关必须为不同算法准备不同的解释方法。^[88]除此之外,个人通过第三方救济机制要求行政机关消除错误信息、撤销错误评级、修复受损信用以及给予相应赔偿的权利也是必不可少的。

其二,由国家基于个人信息权益的保护义务规制信用算法及相应失信惩戒。面对基于数字化权力的失信惩戒,不仅需要建立个人信息权益制度,更需要国家建立强大的数字能力规制算法权力。因为个人往往缺乏足够的力量来预测、判断、控制信息处理行为可能带来的损害。^[89]有论者即指出:“隐藏起来不被数字系统监视或是找出隐藏者,早已成为商业化行为,只有有权的或有钱的才玩得起这个游戏。”^[90]在个体依托个人信息权益无法有效对抗算法权力的情况下,国家必须建立更为强大的作为公共基础设施的公共算法、算力体制以打破算法支配,为人们的数字化生存保驾护航。这也是国家对个人信息受保护权所承担的义务的应有之义。除了建立专责机关根据风险管控原则评判、监管算法行政外,^[91]还可以采取如下规制措施:

一是基于市场化逻辑的算法规制机制。基于算法的高科技属性,行政机关往往要通过与大数据公司的合作才能开发出信用算法。为此,国家应当允许并鼓励多个信用算法公司存在,允许他们提供不同的算法。这样既可以通过市场竞争机制推动信用算法的完善,也为公民通过数据可携带权寻求别的算法评级而避免独家信用算法的垄断性压制提供可能。此外,国家还应当建立官方信用算法机构,为市场主体提供信用评级服务,避免私人公司合谋垄断信用算法,避免商业公司开发的算法自说自话却不受制约。^[92]

二是信用算法的试点试验机制,即通过同一领域、同一类型信用风险的多类型算法试验,来评判哪种信用算法更具包容性、可解释性、准确性,能更好地实现个体发展与风险管控之间的平衡。算法行政不可避免,我们所要做的是建构更为科学的算法,而不是将算法拒之门外。

三是信用算法的算法规制机制。这是至关重要的规制方法。当行政机关基于大数据和算法来评估、预测相对人失信的可能性时,也必须有一个算法机制反向监督信用算法。这个机制需要收集、核实信用算法机制运作过程中的各种数据,并基于大数据方法评估信用算法的效果,为信用算法的监督与改善提供基础。一方面,信用算法不仅是国家治理社会、规训公民的新工具,也是监督政府与平台的工具。这也是《纲要》将政府自身纳入其规范范围,使执法者本身也受到监督的应有之义。如此一来,公民通过向这一系统提供各种异议数据,反过来成为信用算法的监督者,使政府和平台也受制于算法,从而消解了有权者和无权者的界限。^[93]换言之,被信用算法“困住”的,不能仅仅只是被算法评估的“相对人”,还必须包括算法适用者本身。域外即有论者指出,大数据监管的洞察力不必局限于针对犯罪活动,大数据预测分析的力量还可以用于识别警察的不当行为,新的监控技术为观察、监控甚至预测警察的不当行为提供了新的途径。^[94]波斯纳也指出,通过数据挖掘技术,可以让法官发现并反思自己在什么时候会形成偏见。^[95]另一方面,信用算法本质上是一种抽象化处理机制,在追求效率的情况下,算法很难还原具体场景的信息,往往只能根据过去积累的数据作出默认选择。^[96]这就留下了现实与算法结论之间不一致的可能性。因此,建立相应的以算法规制算法的机

[87] Frederick Schauer, *Transparency in Three Dimensions*, 4 *University of Illinois Law Review* 1339, 1351 (2011).

[88] 刘东亮:《新一代法律智能系统的逻辑推理和论证说理》,载《中国法学》2022年第3期,第161—162页。

[89] 王锡铨:《个人信息权益的三层构造及保护机制》,载《现代法学》2021年第5期,第111页。

[90] Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* 55 (Harvard University Press 2015).

[91] 高秦伟:《个人信息的专责机关》,载《法学评论》2021年第6期,第106页。Rory Van Loo, *Rise of the Digital Regulator*, 66 *Duke Law Journal* 1267, 1274 (2017).

[92] [加]雷蒙·安德森:《信用评分工具:自动化信用管理的理论与实践》,李志勇译,中国金融出版社2017年版,第276页。

[93] 虞青松:《算法行政:社会信用体系治理范式及其法治化》,载《法学论坛》2020年第2期,第41页。

[94] Andrew Guthrie Ferguson, *The Rise of Big Data Policing Surveillance, Race, and the Future of Law Enforcement* 5 (New York University Press 2017).

[95] [美]卢克·多梅尔:《算法时代:新经济的新引擎》,胡小锐、钟毅译,中信出版集团2016年版,第144页。

[96] 刘晗、叶开儒:《平台视角中的社会信用治理及其法律规制》,载《法学论坛》2020年第2期,第67页。

制,不仅能满足当事人的异议权,还能通过数据的更新与丰富促成算法机制的改善与革新。

四、结语:迈向大数据时代的法治化失信惩戒

在一定程度上,将失信惩戒视为行政处罚所面临的种种困境,鲜明体现了当下行政法的危机。这些困境表现在,既有行政法框架既无法有效合法化、法治化失信惩戒,又不能匹配其失信治理逻辑,还难以满足其作为新型社会治理手段的体系化要求。在美国学者看来,这一危机根源于经济发展模式的变化,根源于围绕着工业化时代的监管问题和监管能力而设计的行政法未能及时向信息化时代转向。^[97]就失信惩戒法治化而言,这些困境正是因为当下失信惩戒立法者未能认识到:失信惩戒是建立在信用信息与信用算法之上的信用风险管控机制,而不是违法、违约行为之上的制裁机制;违法、违约在失信惩戒中只具有信用信息意义,而无行为意义;当下信用治理大行其道的根本是蓬勃发展的大数据与人工智能突破了政府对数据的收集与处理瓶颈及政府对新型治理机制的需求,而不是因为立法者主观上将其作为制裁手段。而从制裁失信行为到管控交易风险,从规制失信惩戒行为到规制信用算法,失信惩戒法治化进路与中心的转变,既意味着大数据时代法律、法治的变革与发展,也意味着行政法摆脱危机、重新出发的进路。当然,相较于行政处罚式法治化进路,将之理解为风险管控手段进而法治化的难度更大。但是,我们已经迈入大数据时代的这一现实决定了,只有如此推进失信惩戒法治化,才符合社会治理创新的需要与治理革新的趋势,才能真正发挥失信惩戒的社会治理效能。一言以蔽之,只能在发展中解决失信惩戒的法治化问题。

The New Path to Legalize Credit Punishment in Digital Age: From Punishing Trust-Breaking to Risk Management and Control

Chen Guodong

Abstracts: Incorporating credit punishment within the system of administrative penalty is currently a mainstream approach towards the legalization of credit punishment. However, such an approach fails to address the challenge of legitimizing the issue of treating breach of contract and illegal behavior as a breach of trust and subsequently subjecting them to punishment through public law. Additionally, it offers scant relief to interested parties and as such, cannot adequately fulfill the expectation of the social credit system. These constraints also hinder the establishment of the social credit system. From the perspective of controlling resource allocation risks, credit is a vital tool for assessing the credibility of commercial transactions based on credit data. Notably, credit punishment seeks to control such risks by attempting to deter future instances of credit breaches, rather than to punish dishonest behaviors. Implementing credit punishment with a focus on risk management and control conforms to the identity of the administrative agency as a crucial resource allocation subject. Moreover, it meets the needs and the trend of social governance innovation in the era of big data. In this regard, the construction of a legal framework for credit punishment should focus on the credit algorithm while being guided by the principle of risk control. Additionally, it should be attentive to personal information rights and interests and corresponding national protection obligations.

Keywords: credit punishment; legalization; risk management and control; big data; social governance innovation; credit algorithm

(责任编辑:刘馨)

[97] Julie E. Cohen, *The Regulatory State in the Information Age*, 17 *Theoretical Inquiries in Law* 369, 370 (2016).