超越“确保摧毁”：核武器数量、承诺可信度与核威慑原理*

杨 原

【内容提要】 现有几乎所有关于核威慑的理论和政策研究都认为，核武器数量是否达到确保摧毁门槛对核威慑有效性有决定性影响。这种“确保摧毁”原则对学界的影响已达到范式级程度，以至于学者们在争论中国核武器数量够不够这个政策问题时，均自发地从该原则出发论证各自的主张。本文认为，“确保摧毁”范式忽视了承诺可信度在核威慑生效过程中所起的关键性作用，特别是未能充分意识到核武器数量变化对承诺可信度的反向影响以及这种反向影响对核威慑生效机制的影响，因而错将是否具备确保摧毁能力作为判断核威慑是否有效的唯一标准。通过剖析核武器数量、承诺可信度与核威慑有效性三者的作用关系，充分说明，“确保摧毁”既不是核威慑有效的充分条件，也不是其必要条件；未达到确保摧毁门槛时无需增加核武器同样有可能实现有效核威慑，而达到该门槛后为确保威慑有效反而应当（继续）增加核武器。超越“确保摧毁”范式有助于更准确、更完整地理解核威慑生效的全部机制。

【关键词】 核边缘；核威慑；核战略；确保摧毁；战略稳定性

【作者简介】 杨原，中国社会科学院大学国际关系学院副教授，中国社会科学院世界经济与政治研究所副研究员（北京 邮编：100732）。

【DOI】 10.14093/j.cnki.cn10-1132/d.2021.05.001

【中图分类号】 D815.5; E817 【文献标识码】 A 【文章编号】 2095-574X（2021）05-0003-36

* 感谢胡高辰、李彬、吴日强等学者及《国际安全研究》杂志匿名审稿专家的意见和建议。文责自负。

一　导论

本文探讨的问题是，核武器数量与核威慑有效性之间有着什么样的关系。

（一）问题辨析与论文主要观点

核武器数量与核威慑有效性之间有何关系？这个问题与“多少枚核武器才能实现有效核威慑”这个经典问题存在微妙但重要的区别。[①] 后一个问题有一个先验预设：存在一个核武器数量门槛，没有达到这个门槛无法实现有效威慑。同时，后一个问题只提示核武器数量达到多少时有可能实现核威慑，至于是否一定能实现核威慑、如何才能确保实现核威慑，不属于后一个问题的研究范畴。与之相比，本文的研究问题不预先假定存在上述门槛，研究本问题的目的之一就是确定是否存在这个门槛。此外，本文的研究问题不仅关心多少核武器才能满足实现有效核威慑的必要条件（如果不是充分条件的话），更关心不同数量水平下核武器实现威慑功能的具体原理和所需条件。

本文的研究问题源于对“确保摧毁”原则的质疑。现有几乎所有关于核威慑的理论和政策研究都认为，核武器数量是否达到确保摧毁门槛，对核威慑有效性具有决定性影响：确保摧毁不是实现有效核威慑的充分条件，就是实现有效核威慑的必要条件。该原则对学界的影响已达到范式级程度，以至于学者们在争论中国核武器数量够不够这个政策问题时，尽管观点各异，但均自发地从“确保摧毁”原则出发论证各自的主张。几乎所有学者都不反对如下观点：如果核武器数量未达到确保摧毁门槛，则无法实现有效威慑，此时增加核武器是必要的；而一旦达到该门槛，继续增加核武器则是无意义的或者不必要的。[②] 本文的目的是从理论上揭示“确保摧毁”范式的缺陷，并以此为突破口，揭示被该范式所遮蔽的不同核武器数量水平下核威慑生效的所有可能的逻辑和机制。

本文认为，“确保摧毁”范式忽视了承诺可信度在核威慑生效过程中所起的关

① 有不少文献直接以这个经典问题为标题，例如 Alain C. Enthoven and K. Wayne Smith, *How Much Is Enough? Shaping the Defense Program, 1961-1969*, Santa Monica: RAND Corporation, 2005; William Burr, ed., “‘How Much is Enough?’: The U.S. Navy and ‘Finite Deterrence’,” National Security Archive Electronic Briefing Book, No. 275, May 1, 2009, https://nsarchive2.gwu.edu/nukevault/ebb275/index.htm; Keith B. Payne, “Future of Deterrence: The Art of Defining How Much Is Enough,” *Comparative Strategy*, Vol. 29, No. 3, 2010, pp. 217-222; Keir Lieber and Daryl Press, “How Much is Enough? Testing Theories of Nuclear Deterrence,” manuscript, November 2015, https://politics.virginia.edu/wp-content/uploads/2015/11/Lieber-Press-VISC.pdf。

② 相关理论和政策观点见本文第二部分。

键性作用，特别是未能充分意识到核武器数量变化对承诺可信度的反向影响，以及这种反向影响对核威慑生效机制的影响。这是导致现有研究普遍错将“确保摧毁”作为判断核威慑有效与否唯一标准的根本原因。通过剖析核武器数量、承诺可信度与核威慑有效性三者的作用关系，本文指出，“确保摧毁”既不是核威慑有效的充分条件，也不是其必要条件。核武器数量达到确保摧毁门槛，不意味着核威慑一定有效；未达到该门槛，不意味着核威慑一定无效。不同核武器数量水平下的真实核威慑原理的推论与现有共识恰好相反：如果核武器数量未达到确保摧毁门槛，那么维持该数量水平同样有可能实现有效核威慑；如果已经达到，则反而需要为确保威慑有效而（继续）增加核武器数量。

（二）对所使用概念的说明

威慑是本文讨论的核心概念，它包括一般威慑（general deterrence）和即时威慑（immediate deterrence）两类，前者指在和平状态下一国对可能遭受的进攻进行惩罚的宽泛性威胁，后者指在危机状态下一国在对手已经在考虑并准备发动进攻时向其发出的威胁。[①] 这两类威慑在作用机制上存在深刻区别。[②] 本文仅讨论即时威慑，文中提及“核威慑”时均特指即时核威慑。

这种限定主要是基于现实相关性的考虑。核威慑的根本目的是慑止战争，而战争包括了预谋性战争（premeditated war）和意外战争（inadvertent war）两类：前者爆发前通常没有危机，且至少有一方从一开始就有发动战争的意图；后者一般由危机引发，且战争各方至少在危机初期均无意直接发动战争，是危机的延续和升级使各方逐渐认为为了实现既定目标或避免进一步损失而不得不诉诸战争。[③] 从这个分类不难看出，一般威慑的慑止对象是预谋性战争，即时威慑的慑止对象是意外战争。

1945 年之后，由于国际体系的进化、国家间领土规范的强化、对世界大战的负面历史记忆等诸多因素的共同作用，一个大国在没有发生危机的情况下蓄意向另

① Patrick M. Morgan, *Deterrence: A Conceptual Analysis*, Beverly Hills: Sage Publications, 1977, Chapter 2; Patrick M. Morgan, *Deterrence Now*, Cambridge: Cambridge University Press, 2003, Chapter 3.

② 参见 James D. Fearon, “Signaling versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model,” *The Journal of Conflict Resolution*, Vol. 38, No. 2, 1994, pp. 236-269; James D. Fearon, “Selection Effects and Deterrence,” *International Interactions*, Vol. 28, No. 1, 2002, pp. 5-29。

③ Benjamin Miller, “Polarity, Nuclear Weapons, and Major War,” *Security Studies*, Vol. 3, No. 4, 1994, pp. 601-602.

一个大国直接发动大规模战争已经越来越难以想象。[①] 鉴于核武器对使用者自身的巨大成本以及核禁忌规范的存在，一个有核国家在和平状态下突然向另一个国家发动核打击同样不可想象。[②] 就中美两国而言，在和平状态下美国向中国首先发动大规模核打击，在任何情况下都是极其不可能发生的事件。[③] 即使是宣称中美可能陷入“修昔底德陷阱”的格雷厄姆·艾利森（Graham Allison），也承认和强调中美战争风险“通常始于意外事件或事故”。[④] 换言之，预谋性战争这种战争类型在大国群体中已经（近乎）灭绝，因此对当今时代的大国来说，核威慑（以及其他所有形式的威慑）的主要任务是慑止意外战争，即在出现危机的情况下防止危机升级为大规模常规战争或者核战争，这正是即时威慑的任务。

此外，由于不影响论述的总体逻辑，本文不区分“确保摧毁”（assured destruction）和“确保报复”（assured retaliation），它们在文中均指一国在遭到对手首次核打击后所剩核武器能够给对手造成不可承受的损失。[⑤] 相应地，本文也不区

① Shiping Tang, *The Social Evolution of International Politics*, Oxford: Oxford University Press, 2013; Tanisha M. Fazal, *State Death: The Politics and Geography of Conquest, Occupation, and Annexation*, Princeton University Press, 2007; Sharon Korman, *The Right of Conquest: The Acquisition of Territory by Force in International Law and Practice*, Oxford: Clarendon Press, 1996; John Mueller, *Retreat from Doomsday: the Obsolescence of Major War*, New York: Basic Books, 1989. 论述大国无战争现象及其原因的其他相关文献参见 Gary Goertz, Paul F. Diehl, and Alexandru Balas, *The Puzzle of Peace: The Evolution of Peace in the International System*, New York: Oxford University Press, 2016; Joshua Baron, *Great Power Peace and American Primacy: The Origins and Future of a New International Order*, New York: Palgrave Macmillan, 2014; Christopher J. Fettweis, *Dangerous Times? The International Politics of Great Power Peace*, Washington, D.C.: Georgetown University Press, 2010; John Mueller, “War Has Almost Ceased to Exist: An Assessment,” *Political Science Quarterly*, Vol. 124, No. 2, 2009, pp. 297-321; Raimo Vayrynen, ed., *The Waning of Major War*, London: Routledge, 2006。

② 关于核武器对使用者自身的成本，参见 Todd S. Sechser and Matthew Fuhrmann, *Nuclear Weapons and Coercive Diplomacy*, Cambridge: Cambridge University Press, 2017, pp. 13-14；关于核禁忌，参见 Frank Sauer, *Atomic Anxiety Deterrence, Taboo and the Non-Use of U.S. Nuclear Weapons*, New York: Palgrave Macmillan, 2015; Nina Tannenwald, *The Nuclear Taboo: The United States and the Non-Use of Nuclear Weapons Since 1945*, Cambridge: Cambridge University Press, 2007。

③ Thomas J. Christensen, “The Meaning of the Nuclear Evolution: China’s Strategic Modernization and US-China Security Relations,” *Journal of Strategic Studies*, Vol. 35, No. 4, 2012, pp. 448-449.

④ [美] 格雷厄姆·艾利森：《注定一战：中美能避免修昔底德陷阱吗？》，陈定定、傅强译，上海：上海人民出版社 2019 年版，第 219-248 页。

⑤ 有学者指出“确保摧毁”和“确保报复”存在微妙区别，参见 Stephen J. Cimbala, “Deterrence Stability with Smaller Forces: Prospects and Problems,” *Journal of Peace Research*, Vol. 32, No. 1, 1995, pp. 67-68。

分“二次打击能力”和“核报复能力”，它们在文中均指确保摧毁（或确保报复）的能力。最后，考虑到美国是中国主要（甚至唯一）核威慑对象国，而美国核武器对包括中国在内的任何有核国家都能“确保摧毁”，这意味着如果中国核武器数量达到了确保摧毁美国的标准，那么中美之间就自动构成了“相互确保摧毁”，[①] 因此在本文中“确保摧毁”和“相互确保摧毁”所指涉的实际情形是相同的。为叙述简便起见，笔者有时会混用这两个概念。

本文以下分为五个部分：第二部分简要梳理学界对核武器数量与核威慑关系的现有认识。第三和第四部分分别分析核武器数量达到和未达到确保摧毁门槛时的核威慑原理，指出是否达到确保摧毁门槛与能否形成有效威慑无关，但会影响威慑承诺的可信度。第五部分进一步讨论不同核武器数量水平下的核威慑生效机制。最后是结论。

二 核武器数量与核威慑关系的现有认识

学界关于核武器数量与核威慑有效性关系的现有主流认识几乎压倒性地受到“确保摧毁”范式的影响。

这种影响的一个突出表现是，许多学者将核武器数量达到确保摧毁门槛视为实现有效核威慑的充分条件。例如，肯尼思·华尔兹（Kenneth N. Waltz）认为，在一个深度相互确保摧毁的世界，战略军备竞赛将变得没有意义，军事平衡将从根本上保持稳定。[②] 罗伯特·杰维斯（Robert Jervis）同样认为，拥有能够实施报复的核武器就足以确保慑止核进攻和大规模常规军事进攻。[③] 查尔斯·格拉泽（Charles Glaser）则提出，“核武器从根本上消除了安全困境”，这意味着，即使是“一场激烈的超级大国军备竞赛，也几乎不会造成不安全”；对核报复的担忧“还可能降低（任何此类）军备竞赛的强度，从而进一步降低这种不安全”。[④]

认为由核武器所建立起的相互确保摧毁态势是避免核战争乃至常规战争的充

① 这里假定中美两国在中国是否达到确保摧毁美国的门槛上的认知不存在分歧。

② Scott D. Sagan and Kenneth N. Waltz, eds., *The Spread of Nuclear Weapons: A Debate*, New York: W. W. Norton, 1995, pp. 30-31.

③ Robert Jervis, *The Meaning of the Nuclear Revolution: Statecraft and the Prospect of Armageddon*, Ithaca, NY: Cornell University Press, 1989, Chapter 1.

④ Charles L. Glaser, *Analyzing Strategic Nuclear Policy*, Princeton, NJ: Princeton University Press, 1990, pp. 95, 97.

分条件的最具代表性的观点，来自冷战时期麦乔治·邦迪（McGeorge Bundy）提出的“存在性威慑”（existential deterrence）概念。[①] 邦迪认为，存在性威慑与那些基于战略理论或者宣示性政策甚至需要依赖于国际承诺的威慑不同，其生效只依赖于一个条件，那就是各方都拥有大量的可用于对付对手的热核武器。[②] 当前很多政策分析将这种存在即威慑的理念进一步极端化，认为一旦一国拥有了核武器，就自动拥有了威慑能力。[③]

上述思想也在很大程度上影响了中国学界对有效核威慑所需条件的认识。例如，《中国青年报》2018 年的一篇评论称：“一个国家只要具有了可靠的二次打击能力，那么它就具备了有效的核威慑能力。”[④] 李彬和胡高辰在一篇探讨中国核威慑有效性的论文中表示，“对中国核威慑能力有效性的判断，可以用中国核报复能力来衡量”，并进一步解释称，“核报复能力指一个国家在遭受了对手的第一波核打击后，穿透对手导弹防御系统、向对手进行核反击的能力。如果一个国家的核报复能力能够让对手产生不可承受的损失，对手就不敢轻易对该国发动先发制人的核打击”。[⑤]

其他一些学者尽管否认确保摧毁是核威慑生效的充分条件，但认为前者是后者的必要条件。例如，在最早提出战略稳定性概念的《精致的恐怖平衡》一文中，艾尔弗雷德·沃尔斯泰特（Alfred Wohlstetter）就指出，可靠的二次打击能力是战略稳定性——双方都没有动机挑起武装冲突的状态——存在的重要条件。[⑥] 中国学者鹿音明确表示，“一个国家只有在具备了核报复能力后，才能有效慑止对手的核攻

① McGeorge Bundy, “To Cap the Volcano,” *Foreign Affairs*, Vol. 48, No. 1, 1969, pp. 1-20; McGeorge Bundy, “Existential Deterrence and Its Consequences,” in Douglas MacLean, ed., *The Security Gamble: Deterrence Dilemmas in the Nuclear Age*, Totowa, NJ: Rowman and Allanheld, 1984, pp. 3-13.

② McGeorge Bundy, “Existential Deterrence and Its Consequences,” in Douglas MacLean, ed., *The Security Gamble: Deterrence Dilemmas in the Nuclear Age*, Totowa, NJ: Rowman and Allanheld, 1984, pp. 8-9.

③ 参见 Matthew H. Kroenig, “Time to Attack Iran: Why a Strike Is the Least Bad Option,” *Foreign Affairs*, Vol. 91, No. 1, 2012, pp. 76-86; Jon Wolfsthal, “How to Reason With a Nuclear Rogue,” *Foreign Policy*, July 12, 2017, http://foreignpolicy.com/2017/07/12/north-korea-nukes-icbm-test-nuclear-weapons/。

④ 东方：《增加战略核武器数量加强核力量不符合世界潮流》，载《中国青年报》2018 年 8 月 1 日，第 6 版。

⑤ 李彬、胡高辰：《美国视阈中的中国核威慑有效性》，载《外交评论》2018 年第 5 期，第 23 页。

⑥ Albert Wohlstetter, ‘‘The Delicate Balance of Terror,’’ *Foreign Affairs*, Vol. 37, No. 2, 1959, pp. 211-234.

击”。[①] 类似地，另一位中国学者刘冲认为，“按照核威慑战略，失去核反击能力就意味着核威慑无效”。[②]

无论认为确保摧毁是实现核威慑的充分条件还是必要条件，学界各派均认同确保摧毁对核威慑有效性有决定性影响，并（有意或无意地）将能否达到确保摧毁门槛作为判断一国是否拥有有效核威慑能力的事实上的唯一标准。例如，作为否认中国具有对美核威慑能力的代表性学者，基尔·利伯（Keir A. Lieber）和达里尔·普雷斯（Daryl Grayson Press）反复论证其观点的核心理由非常简单直白，就是（他们认为）中国核武器在美国首次核打击下生存能力弱、不具备二次打击能力。[③] 而对这种观点的反驳，即认为中国具备对美核威慑能力，也同样是从报复能力方面来论证的。例如金骏远（Avery Goldstein）、李彬、吴日强等学者从不确定性角度，强调美国很难保证在首次打击中百分之百地摧毁中国所有的核武器，因此中国仍然有可能有足够数量的核武器存活下来对美国进行报复，并制造不可承受的损失。[④] 其他承认中国核威慑能力的学者，也几乎都是从中国核武器的生存能力、机动性、隐蔽性等有助于保证二次打击能力的因素加以论证的。[⑤]

在近期国内出现的关于中国是否应当增加核武器数量的政策争论中，争论各方尽管观点各异，但都接受和强调“确保摧毁”原则。一位反对中国增加核武器数量的学者认为，中国的核战略是“精干有效”，其中“有效”指的是“在不首先使用核武器的前提下，能扛得住敌人的第一波核打击，还能有效进行反击”：假如美国

① 鹿音：《中美战略稳定关系的演进》，载《当代美国评论》2017 年第 2 期，第 25 页。

② 刘冲：《核武器与常规军事冲突的关系》，载李彬、赵通主编：《理解中国核思维》，北京：社会科学文献出版社 2016 年版，第 115 页。

③ Keir A. Lieber and Daryl Grayson Press, “The End of MAD? The Nuclear Dimension of U. S. Primacy,” *International Security*, Vol. 30, No. 4, 2006, pp. 7-44; Keir A. Lieber and Daryl Grayson Press, “Superiority Complex: Why America’s Growing Nuclear Supremacy May Make War with China More Likely,” *The Atlantic Monthly*, July/August 2007, pp. 86-92; Keir A. Lieber and Daryl Grayson Press, “U. S. Nuclear Primacy and the Future of the Chinese Deterrent,” *China Security*, Winter 2007, pp. 66-89.

④ Avery Goldstein, *Deterrence and Security in the 21st Century: China, Britain, France, and the Enduring Legacy of the Nuclear Revolution*, Stanford: Stanford University Press, 2000, pp. 121-136; Li Bin, “Paper Tiger with Whitened Teeth,” *China Security*, Autumn 2006, pp. 78-89; Wu Riqiang, “Certainty of Uncertainty: Nuclear Strategy with Chinese Characteristics,” *Journal of Strategic Studies*, Vol. 36, No. 4, 2013, pp. 579-614; Wu Riqiang, “Living with Uncertainty: Modeling China’s Nuclear Survivability,” *International Security*, Vol. 44, No. 4, 2020, pp. 84-118.

⑤ 相关梳理和文献参见李彬、胡高辰：《美国视阈中的中国核威慑有效性》，载《外交评论》2018 年第 5 期，第 37-38 页。

对中国发起核打击，虽然中国核武器数量比美国少，但已有的核武器能够在反击中把美国“打亡国”，就足够了。[①] 另一篇反对增加核武器的评论同样认为，中国现有的核威慑战略得以成立的一个关键点，是“中国需要确保在经历对方核打击之后还可以拥有足够令对手无法承受的核反击力量”。[②]

主张中国增加核武器的观点同样是基于确保摧毁逻辑，认为按照罗伯特·麦克纳马拉（Robert S. McNamara）的“不可承受”标准，确保摧毁所需核武器规模应大于 400 枚 1000kt 弹头或 600 枚 650kt 弹头，而中国现有核武器数量很可能达不到这个标准。[③] 主张改进中国核武器质量以提高生存和突防能力的观点同样提出，决定中国核武器数量是否足够的首要标准，是能否“有效地发起二次核反击并保证效果”，[④] 而根据这个标准，中国的核报复能力还不够显著。[⑤] 综上可见，争论各方在是否应当增加核武器这个问题上存在一个原则性共识：如果核武器数量未达到确保摧毁门槛，则无法实现有效威慑，此时增加核武器是必要的；而一旦达到该门槛，继续增加核武器则是无意义的或者不必要的。

本文以下两部分将指出，以上认识夸大了“确保摧毁”对核威慑的意义，忽视了影响核威慑有效性的其他因素和可能机制。

三　核武器数量达到确保摧毁门槛时的核威慑原理

“达到确保摧毁门槛是实现有效核威慑的关键”这种认识的错误首先在于，（相互）确保摧毁不是实现有效核威慑的充分条件。即使一国的核武器数量已经确定无疑地超过了确保摧毁的门槛并且敌国对此深信不疑，依然无法保证核威慑的有效性。

（一）威慑的本质是承诺

“承诺”（commitment）是指承诺方向被承诺方就做某事或不做某事所做的一种

① 李司坤、刘煊尊：《中国扩充核武有无必要？》，环球网，2020 年 5 月 9 日，https://world.huanqiu.com/article/ 3yA8uuetxLm。

②《为何少量核武器也能维持和平？中国的核战略逻辑比较复杂》，今日头条，2020 年 5 月 10 日，https://www.toutiao.com/a6825186867642630660/。

③《中国究竟需要多少核武器？1000 枚绝不嫌少，扩核势在必行》，今日头条，2020 年 5 月 11 日，https://www.toutiao.com/a6825413069464666636/。

④《中国到底要多少枚核武才够？》，2020 年 5 月 10 日，https://m.k.sohu.com/d/450018974?channelId=1&page=1。

⑤ 吴日强：《中美战略武器技术竞争前瞻》，澎湃新闻，2020 年 4 月 30 日，https://www.thepaper.cn/newsDetail_forward_7205991。

保证，它包括了“许诺”（promise）和“威胁”（threat）两种情况，前者是奖励性保证，后者是惩罚性保证。[①] 本文所讨论的核威慑就是一种承诺，更具体地说是一种威胁。A 国向 B 国实施核威慑，本质上是 A 国向 B 国做出的这样一种威胁：危机状态下如果 B 国（用核武器或常规武器）进攻 A 国，那么 A 国保证会用核武器对 B 国进行报复并使 B 国遭受不可承受的损失。

A 国对 B 国的核威慑能否生效，取决于 A 国上述威胁的可信度（credibility）。威胁的可信度包括了两个维度：威胁能力的可信度和威胁意图的可信度。两者缺一不可。A 国威胁能力的可信度是指 B 国在多大程度上相信 A 国有对 B 国制造不可承受的损失的核打击能力，[②] 该可信度直接取决于 A 国的硬实力。许多研究在讨论中国核威慑的可信度问题时，实际讨论的是这个维度的威胁可信度；[③] 前文所梳理的各派争论，其争论焦点同样是中国核威胁能力的可信度。

A 国威胁意图的可信度是指 B 国在多大程度上相信 A 国会在相应条件下履行其承诺，即向 B 国实施核报复。因此威胁意图的可信度也常被称为“承诺可信度”。一个承诺是可置信的（credible），是指履行该承诺是承诺者的理性选择，亦即履行该承诺的预期效用比不这样做的预期效用大。[④] 因此，A 国威胁意图的可信度取决于 B 国在多大程度上认为 A 国在相应条件下实施核报复对 A 国自己的效用大于不实施核报复的效用。如果 B 国认为 A 国实施核报复对 A 国自己来说得不偿失，那么 B 国就不会相信 A 国会履行其核威慑承诺，在这种情况下，A 国对 B 国的核威慑无效。

---

① [美] 托马斯·谢林：《承诺的策略》，王永钦、薛峰译，上海：上海人民出版社 2009 年版，第 1 章；James W. Davis, Jr., *Threats and Promises: The Pursuit of International Influence*, Baltimore: The Johns Hopkins University Press, 2000。

②（能够制造不可承受的损失的）核打击能力当然是核威慑的必要条件，但如本文第四部分将要论述的，二次核打击能力（亦即核报复能力）不是必要条件。

③ 参见 Baohui Zhang, “US Missile Defence and China’s Nuclear Posture: Changing Dynamics of an Offence–Defence Arms Race,” *International Affairs*, Vol. 87, No. 3, 2011, pp. 555-569; Michael S. Chase, “China’s Transition to a More Credible Nuclear Deterrent: Implications and Challenges for the United States,” *Asia Policy*, Vol. 16, No. 1, 2013, pp. 69-101。

④ Richard Ned Lebow, *Between Peace and War: The Nature of International Crisis*, Baltimore, MD: Johns Hopkins University Press, 1981, p. 15; Richard K. Betts, *Nuclear Blackmail and Nuclear Balance*, Washington, DC: Brookings Institution, 1987, p. 12; Robert Powell, “Crisis Bargaining, Escalation, and MAD,” *The American Political Science Review*, Vol. 81, No. 3, 1987, p. 719; D. Marc Kilgour and Frank C. Zagare, “Credibility, Uncertainty and Deterrence,” *American Journal of Political Science*, Vol. 35, No. 2, 1991, pp. 306-308.

（二）确保摧毁会导致威慑承诺不可信

显然，“确保摧毁”解决的是威胁能力的可信度问题，但它解决不了威胁意图的可信度问题。不仅如此，由“确保摧毁”能力所建立起的“相互确保摧毁”态势还会削弱威胁意图可信度。这是因为，对对方进行的核报复同样有可能招致对方的反报复，在“相互确保摧毁”态势下，这种反报复将“确保”己方被摧毁。

威胁能力可信度与威胁意图可信度之间的这种张力其实已为一些学者所注意。丹尼尔·埃尔斯伯格（Daniel Ellsberg）早在 1959 年就提出了核威慑的一个根本难题：在实施威胁的成本令人望而却步的情况下，该如何通过威胁来阻止挑战者？[①] 巴里·内勒巴夫（Barry Nalebuff）甚至因此认为，核威慑的存在并不是建立在理性计算的基础上的，因为核战争会摧毁我们试图拯救的东西，因此以核战争相威胁是不理性的，因而也是不可置信的。[②] 帕特里克·摩根（Patrick M. Morgan）则将这种张力视为威慑理论的一个重要的矛盾之处（inconsistency）：在相互威慑关系中，进攻方（即被威慑方）同样能够对威慑方造成不可承受的损失，因此，威慑方对进攻方的进攻进行报复所面临的自身损害至少不会比不进行报复所遭受的损害少，在这种情况下，进攻方很难相信威慑方会实施这种报复。[③] 托德·塞克瑟尔（Todd S. Sechser）和马修·富尔曼（Matthew Fuhrmann）同样指出，实施核打击会使实施方自己付出很高的代价，因此被威胁方并不一定会相信这种核威胁会真的实施。[④]

严格的形式模型分析支持学者们对核威慑可靠性的这种担忧。有研究甚至认为，决定威慑稳定性的唯一关键变量是威慑方发出的威胁的初始可信度（即威慑方实施该威胁对自身的效用是否以及多大程度上优于不实施该威胁而听任被威慑方实施其挑战对前者的效用）。冲突成本的提高并不必然增加威慑的稳定性，相反，超过某一临界点后，冲突成本的增加对于威慑效果而言要么是不必要的，要么是适得其反的。对于那些决心不足的威慑者来说，冲突的成本越大，越有可能选择退让

① Daniel Ellsberg, “The Theory and Practice of Blackmail,” Lecture at the Lowell Institute, Boston, MA, March 10, 1959, reprinted in Oran R. Young, ed., *Bargaining: Formal Theories of Negotiation*, Urbana: University of Illinois Press, 1975, pp. 358-359, quoted from Frank C. Zagare and D. Marc Kilgour, “Asymmetric Deterrence,” *International Studies Quarterly*, Vol. 37, No. 1, 1993, p. 2.

② Barry Nalebuff, “Brinkmanship and Nuclear Deterrence: The Neutrality of Escalation,” *Conflict Management and Peace Science*, Vol. 9, No. 2, 1986, p. 19.

③ Patrick M. Morgan, *Deterrence Now*, Cambridge: Cambridge University Press, 2003, p. 47.

④ Todd S. Sechser and Matthew Fuhrmann, “Crisis Bargaining and Nuclear Blackmail,” *International Organization*, Vol. 67, No. 1, 2013, pp. 173-195.

从而导致威慑的失败。[①] 从这个意义上讲，核武器的巨大破坏力反而是导致核威慑承诺可信度降低、核威慑失败风险增加的内生性原因。

正是由于“相互确保摧毁”态势下核威慑承诺不可信，才使得历史上那些以“确保摧毁”为理论支撑的核威慑政策不断遭到质疑。罗伯特·鲍威尔（Robert Powell）指出，大规模报复战略之所以不可行，根本原因就是美苏之间建立了“相互确保摧毁”。在“相互确保摧毁”状态下，美国以核报复相威胁以保护哪怕是美国的关键利益也是难以令苏联置信的，因为美国的核报复会导致苏联毁灭性的反报复，这会使得美国实施其威胁的代价比放任苏联所为的代价更大。[②] 基于同样的逻辑，托马斯·谢林（Thomas C. Schelling）更具体地指出，1962 年 10 月 22 日，美国总统约翰·肯尼迪（John F. Kennedy）宣布“任何从古巴发射的针对西半球任何国家的核导弹”都将导致“对苏联的全面报复”，这一政策是“非理性的”。[③]

（三）确保摧毁不是“以核慑常”成功的充分条件

从承诺可信度的角度不难看出，可靠的二次核打击力量不是“以核慑常”成功的充分条件。正如利伯和普雷斯所指出的，如果进攻方能够以同样的方式进行反报复，那么防御方对进攻方的常规军事进攻发出核报复的威胁就很难令进攻方置信。[④] 实证研究的结果也支持这一论断。马克·贝尔（Mark S. Bell）和尼古拉斯·米勒（Nicholas L. Miller）在对此前的统计研究进行改进后发现，两个拥有核武器的国家发生战争的可能性并不会显著降低。[⑤] 维平·纳兰（Vipin Narang）更进一步地通过统计研究证明，一个国家即使获得了确保摧毁的核能力，它所面临的武装冲突——哪怕是严重的武装冲突——的风险也不会消失；确保报复策略几乎没有系统性的慑止冲突爆发和升级的作用，它甚至无法慑止全面战争。[⑥] 从经验上看，

① Frank C. Zagare and D. Marc Kilgour, “Asymmetric Deterrence,” *International Studies Quarterly*, Vol. 37, No. 1, 1993, pp. 1-27, especially at p. 19.

② Robert Powell, “Crisis Bargaining, Escalation, and MAD,” *The American Political Science Review*, Vol. 81, No. 3, 1987, p. 718.

③ Thomas C. Schelling, *Arms and Influence*, New Haven: Yale University Press, 1966, pp. 40-43.

④ Keir A. Lieber and Daryl G. Press, *The Myth of the Nuclear Revolution: Power Politics in the Atomic Age*, Ithaca: Cornell University Press, 2020, p. 8 and Chapter 4.

⑤ Mark S. Bell and Nicholas L. Miller, “Questioning the Effect of Nuclear Weapons on Conflict,” *The Journal of Conflict Resolution*, Vol. 59, No. 1, 2015, pp. 74-92.

⑥ Vipin Narang, “What Does It Take to Deter? Regional Power Nuclear Postures and International Conflict,” *The Journal of Conflict Resolution*, Vol. 57, No. 3, 2012, pp. 494-501; Vipin Narang, *Nuclear Strategy in the Modern Era: Regional Powers and International Conflict*, Princeton: Princeton University Press, 2014, Chapter 9.

在历史上有核国家之间发生过的总共两次战争（或准战争）中，战争发起方所面对的对手都采取的是确保报复策略。[①]

（四）确保摧毁不是“以核慑核”成功的充分条件

首先，可靠的二次核打击力量难以确保慑止低当量的针对军事力量的核打击。认为相互确保摧毁能够慑止核打击的一个主要理由是，在进攻方对防御方实施首次核打击后，防御方已没有更多可损失的，这使得后者没有理由不进行核报复。然而，正如哈里森·瓦格纳（R. Harrison Wagner）所指出的，进攻方可以通过将防御方的核力量作为首次核打击的目标，亦即通过打击军事力量战略（counterforce strategy），[②] 使防御方的核报复威胁变得不可置信：一方面，针对军事力量的核打击能够减少防御方用于核报复的核武器数量，使得进攻方在遭到防御方核报复后有更大可能剩下足够多的核武器，从而更有把握对防御方的核报复进行反报复；另一方面，更重要的是，进攻方首次核打击不针对防御方的主要城市等战略目标，这显著地限制了首次核打击对防御方造成的损失。[③] 这意味着，在遭到进攻方首次核打击后，理性的防御方将不得不在“退让从而只承受有限损失”和“进行核报复从而招致反报复并由此承受被核摧毁的损失”两个选项中选择前者。[④]

扎卡里·兹瓦尔德（Zachary Zwald）同样指出，如果首次使用核武器是基于理性的决策，进攻方就会将核打击限定为低当量的针对军事力量的打击，在这种情况下，以威胁使用高当量核武器进行报复难以起到威慑作用。这正是确保摧毁理论自20世纪70年代以来一直受到质疑的主要原因。[⑤] 而当前，精确运载系统、新的侦

① Vipin Narang, “What Does It Take to Deter? Regional Power Nuclear Postures and International Conflict,” *The Journal of Conflict Resolution*, Vol. 57, No. 3, 2012, p. 501. 这两次战争（或准战争）分别是1969年中苏边境冲突和1999年印巴卡吉尔战争。

② 打击军事力量战略与损害限制战略（damage limitation）的意图相近，意在通过发动先发制人式打击减少对方能够打击到己方领土的核武器数量，由此限制己方在未来可能发生的全面核战争中的受损程度。参见 Charles L. Glaser and Steve Fetter, “Should the United States Reject MAD? Damage Limitation and U.S. Nuclear Strategy toward China,” *International Security*, Vol. 41, No. 1, 2016, p. 49。

③ 值得注意的是，防御方通过分散、隐蔽和机动等方式减少进攻方首次核打击所消灭的核武器数量，并不妨碍反而会加强这一逻辑。

④ R. Harrison Wagner, “Nuclear Deterrence, Counterforce Strategies, and the Incentive to Strike First,” *American Political Science Review*, Vol. 85, No. 3, 1991, pp. 727-749.

⑤ Zachary Zwald, “Imaginary Nuclear Conflicts: Explaining Deterrence Policy Preference Formation,” *Security Studies*, Vol. 22, No. 4, 2013, p. 653.

察技术以及核武库相对于冷战时期的削减等因素，正在增加对核武库进行常规和核打击的可行性。更重要的是，高精度发射系统与核武器相结合，有可能实现几乎不产生放射性沉降物的目标打击战略，从而大大减少死亡人数。有学者将这种变化称为打击军事力量战略的革命（counterforce revolution）。[①] 在这种技术变革背景下，首次针对军事力量的核打击对防御方造成的损失，与防御方实施核报复后进攻方反报复对防御方造成的损失之间的差异会更显著，这无疑会进一步削弱防御方对这类核打击的威慑承诺可信度。

国防大学战略研究所原所长潘振强少将设想过一个具体情境：美国为阻挠中国统一台湾，使用低当量战术核武器对中国海军舰队发动小型战术核打击，导致中国数千士兵死亡。在这种情况下，中国如果对美国实施惩罚性核报复，根据中国现有的核报复能力，这种报复意味着对美国大城市实施核打击并由此造成数十万乃至数百万人死亡以及其他破坏性影响。对此潘振强坦言，如何使美国相信中国有足够的政治勇气和决心采取这种不成比例的报复，可能是一个大问题。[②] 而比人道主义层面的顾虑更棘手的是，中国决策层还必须考虑实施核报复将极有可能遭到美国的反报复，从而使自己遭受更大的损失。如果美国认为中国不愿意承受这一更大的损失，那么中国对美国这种类型核打击的核威慑就将难以奏效。

其次，“确保摧毁”有可能诱发核升级。获得确保摧毁能力有可能极大地提升本国对危机稳定性的信心，亦即相信敌我两国的常规冲突不会升级为核冲突，而这种信心有时非但无益于慑止敌国的进攻，反而有可能加大双方发生核冲突的几率。菲奥娜·坎宁安（Fiona S. Cunningham）和傅泰林（M. Taylor Fravel）在 2019 年的一篇论文中，就坦率而尖锐地指出了中国所面临的这一风险。他们指出，中国战略界对中美危机稳定性的判断非常乐观，认为在相互确保摧毁态势下，中美常规军事冲突甚至有限核冲突升级为大规模核战争的风险很低。这种乐观预期会被美方解读为，美国对中国发动常规和有限核打击引发中国大规模核报复的可能性很低，这无疑将增加美国发动这两类打击的可能性，这反过来又将增加中国的安全压力，迫使

① Keir A. Lieber and Daryl Grayson Press, “The End of MAD? The Nuclear Dimension of U.S. Primacy,” *International Security*, Vol. 30, No. 4, 2006, pp. 7-44; Keir A. Lieber and Daryl G. Press, “The New Era of Nuclear Weapons, Deterrence, and Conflict,” *Strategic Studies Quarterly*, Vol. 7, No. 1, 2013, pp. 3-14.

② Zhenqiang Pan, “A Study of China’s No-First-Use Policy on Nuclear Weapons,” *Journal for Peace and Nuclear Disarmament*, Vol. 1, No. 1, 2018, p. 129.

中国实施报复以阻止美国的进一步打击，进而引发核升级。[①]

再次，“相互确保摧毁”甚至不能确保慑止大规模核战争。如上所述，低当量、针对军事力量的有限核打击制造的损失相对有限，因此在相互确保摧毁态势下，威慑方以自己被摧毁为代价实施针对这类核打击的核报复的承诺不可信。可如果敌国意图发动的是大规模无限制核打击，“确保摧毁”能力能够确保避免这个最危险的情景变为现实吗？答案很可能也是否定的。尽管人类迄今尚未发生过核战争，但这并不意味着我们对核安全机制的认识已经尽善尽美，没有爆发核战争有可能仅仅是一种幸运。[②] 事实上，将冷战时期美苏没有发生直接战争完全归因于核威慑缺乏足够的实证根据。[③] 科林·格雷（Colin S. Gray）也说：“冷战期间不可能有能够打赢的核战争并不是一种权威的真理……任何结果都有可能，取决于当时的特殊情况。”[④]

迄今为止，人类尚未发生过核战争，这是人类的大幸运，不过也是本文的小“不幸”：我们因此缺乏“以核慑核”失败的真实案例以证明确保摧毁不是慑止大规模核战争的充分条件。为克服这一困难，同时也为了使我们能够更直观地理解相互确保摧毁态势下核威慑失败的核心机理，这里不妨援引科幻小说《三体》中的一个与核威慑失败相近似的虚拟案例。科幻小说当然是虚构的，但某些基于自然科学或社会科学原理的科幻小说情节类似“思想实验”，能够为我们思考极端条件下的某些自然或社会过程提供启发。事实上，借助科幻作品研究国际关系的严肃学术专著早有先例。[⑤]

在《三体》中，来自三体星系的高智慧文明意图入侵地球，人类与三体人的科技水平差距悬殊，无力抵御。关键时刻，小说主人公罗辑发现了宇宙的“黑暗森林”法则，其核心思想是，任何暴露自己宇宙位置的文明都将遭到其他更高级文明的毁

① Fiona S. Cunningham and M. Taylor Fravel, “Dangerous Confidence? Chinese Views on Nuclear Escalation,” *International Security*, Vol. 44, No. 2, 2019, p. 105.

② Erik Gartzke and Matthew Kroenig, “Nukes with Numbers: Empirical Research on the Consequences of Nuclear Weapons for International Conflict,” *Annual Review of Political Science*, Vol. 19, 2016, p. 398.

③ 李彬：《核武器的强制性作用》，载张沱生主编：《核战略比较研究》，北京：社会科学文献出版社 2014 年版，第 391 页。

④ Colin S. Gray, *Modern Strategy*, New York: Oxford University Press, 1999, p. 314.

⑤ 例如 Stephen Benedict Dyson, *Otherworldly Politics: The International Relations of Star Trek, Game of Thrones, and Battlestar Galactica*, Baltimore: Johns Hopkins University Press, 2015。

灭性打击。[①] 罗辑利用这一法则向三体人承诺，如果后者不放弃入侵，他将向宇宙发布三体星系坐标，这将导致人类与三体人的共同毁灭。人类由此与三体人建立了“相互确保摧毁”关系。罗辑向三体人发出上述威胁时，他由于人类对他的误解已决意自尽，在这种情况下，他的威胁成功慑止了三体人的入侵。然而，这种威慑只维持了 54 年，当负责在遭到入侵时按下发布星系位置按钮的“执剑人”由罗辑换成另一个心地慈善的女孩程心的那一刻，三体人发动了对地球的进攻。由“黑暗森林”法则所确立的“相互确保摧毁”态势并未改变，但威慑失效了，因为三体人根据获得的信息判断，程心没有不惜牺牲全人类而按下按钮的决心。[②]

这个虚拟案例形象地说明了“相互确保摧毁”为何无法确保慑止无限战争：当你的报复会导致你自己被毁灭时，让对手相信你会实施这种报复是困难的，而如果对手不相信你会报复，那么你纵然有百分之百摧毁对手的报复能力，这种能力也将形同虚设。

（五）失控风险、决心与相互确保摧毁下的核威慑原理

以上分析了承诺可信度（威胁意图可信度）对核威慑的重要意义，以及“确保摧毁”能力对承诺可信度的妨碍作用。那么，在相互确保摧毁条件下，如何才能克服承诺不可信问题而实现有效核威慑呢？谢林在其名著《冲突的战略》中给出了这个难题的经典解决方案：制造一种不能完全掌控的威胁。[③]

谢林指出，威慑方可以通过故意制造危机甚至发动有限战争，或者以强硬姿态使业已发生的冲突升级，增加事态失控而引发灾难性核冲突的风险，以这种风险慑止被威慑方的进攻。[④] 这种风险当然同样是被威慑方和威慑方都不愿承担的，但与直接实施核报复相比，制造这种风险的每一步行动都并不必然导致全面核冲突，预

① “黑暗森林”法则是《三体》虚构的一个概念，但内涵与政治学中的“霍布斯自然状态”思想以及国际关系理论中的“安全困境”概念相吻合，其主要观点是：所有文明的首要目标是生存，而每个文明都难以确定和相信其他文明的意图。在宇宙总资源有限而每个文明对资源的需求又在不断增加的情况下，每个文明都无法保证其他文明不会在未来某个时刻入侵和消灭自己，因此每个文明的理性选择都应是小心隐藏自己，同时对那些暴露自己位置的文明，应趁其文明程度尚未构成威胁时对其发动预防性打击。参见刘慈欣：《三体 II：黑暗森林》，重庆：重庆出版社 2008 年版，第 441-448 页。

② 刘慈欣：《三体 II：黑暗森林》，重庆：重庆出版社 2008 年版；刘慈欣：《三体 III：死神永生》，重庆：重庆出版社 2010 年版。

③ 原文是“make a threat that leaves something to chance”。这个经典表述被其后历代核威慑理论文献所反复引用。

④ Thomas C. Schelling, *The Strategy of Conflict*, Cambridge, MA: Harvard University Press, 1960, Chapter 8; Thomas C. Schelling, *Arms and Influence*, New Haven: Yale University Press, 1966, pp. 99-125.

期成本远低于直接实施核报复的预期成本，因此更加可信。[①] 以这种方式实现威慑的关键是，威慑方必须设法将事态推向失控的边缘，使冲突的进程和前景在一定程度上超出冲突双方的主观可控范围。[②]

这种威慑方式生效的过程类似下面这种情境：两个人被绳子绑在一起站在悬崖边，显然，其中任何一人对另一人说“你如果攻击我，我就把你推下悬崖”，后者都不会相信。但前者可以故意把两人尽量拖向悬崖边缘，甚至一只脚悬空，从而造成一种随时可能失控的不稳定状态。此时，两人的任何肢体碰撞，或者脚下石子滑落，甚至一阵风，抑或前者恐高引起的眩晕，都有可能使两人跌下悬崖，此时停止攻击、主动后退就会成为后者的理性选择。[③] 这种威慑方式也因此被称为“边缘策略”（brinkmanship）。另一个类似的情境是两个不会游泳的人在一条独木舟上，其中一人为慑止另一人的攻击，显然不能以将舟弄翻相威胁，但前者可以将舟左右摇晃。舟摇晃的幅度越大，两人淹死的风险就越大，但同时，后者为规避这一风险而停止攻击的概率也越大。[④]

从上述情境所展示的威慑生效原理可以看出，在冲突失控将给双方都造成不可承受的损失的情况下，博弈双方谁愿意承担更大的失控风险，谁就能够迫使另一方让步。[⑤] 愿意承担的失控风险大小反映的是在冲突中坚持下去的决心（resolve），

① Robert Powell, “Crisis Bargaining, Escalation, and MAD,” *The American Political Science Review*, Vol. 81, No. 3, 1987, p. 719.

② Thomas C. Schelling, *The Strategy of Conflict*, Cambridge, MA: Harvard University Press, 1960, p. 188.

③ 参见 Thomas C. Schelling, *Arms and Influence*, New Haven: Yale University Press, 1966, p. 99。

④ 参见 Thomas C. Schelling, *The Strategy of Conflict*, Cambridge, MA: Harvard University Press, 1960, p. 196; Thomas C. Schelling, *Arms and Influence*, New Haven: Yale University Press, 1966, p. 91。

⑤ Thomas C. Schelling, *Arms and Influence*, New Haven: Yale University Press, 1966, pp. 92-125; Glenn H. Snyder and Paul Diesing, *Conflict among Nations: Bargaining, Decision Making, and System Structure in International Crises*, Princeton: Princeton University Press, 1977, pp. 456-458; Robert Jervis, “Deterrence Theory Revisited,” *World Politics*, Vol. 31, No. 2, 1979, pp. 291-292; Robert Jervis, “Why Nuclear Superiority Doesn’t Matter,” *Political Science Quarterly*, Vol. 94, No. 4, 1979/1980, p. 631; Kenneth N. Waltz, “Toward Nuclear Peace,” in Robert J. Art and Kenneth N. Waltz, eds., *The Use of Force* (Second Edtion), Lanham: University Press of America, 1983, pp. 589-590; Robert Jervis, *The Illogic of American Nuclear Strategy*, Ithaca: Cornell University Press, 1984, pp. 126-46; Robert Powell, “The Theoretical Foundations of Strategic Nuclear Deterrence,” *Political Science Quarterly*, Vol. 100, No. 1, 1985, p. 78; Richard K. Betts, *Nuclear Blackmail and Nuclear Balance*, Washington D.C.: The Brookings Institution, 1987, pp. 11-12; Robert Jervis, *The Meaning of the Nuclear Revolution*, Ithaca, NY: Cornell University Press, 1989, pp. 38-41, 105; Robert Powell, *Nuclear Deterrence Theory: The Search for Credibility*, Cambridge: Cambridge University Press, 1990, pp. 33-45; Robert Powell, “Nuclear Deterrence Theory, Nuclear Proliferation, and National Missile Defense,” *International Security*, Vol. 27, No. 4, 2003, pp. 89-97.

它取决于双方在冲突中的相对利害（stake）大小。[①] 核边缘态势下利害、决心与威慑结果之间的确切关系可以借助以下模型加以说明。[②]

设有两个国家 I 和 II 处于危机中，其中 I 是潜在进攻方，它首先在升级冲突和让步之间进行选择。如果 I 选择让步，则博弈结束，I 和 II 的支付分别为 s_I 和 w_{II}，其中 s 是向对手让步的支付，w 是在危机中获胜的支付。如果 I 选择升级，则轮到防御方 II 在升级和让步之间做出选择。如果 II 选择让步，则博弈结束，I 和 II 分别获得 w_I 和 d_{II}。如果 II 选择升级，则博弈继续，但双方都将承受一定概率的核升级风险。如果随之发生了核战争，则双方分别获得 d_I 和 d_{II}。如果没有发生核战争，则轮到 I 继续在让步和进一步升级之间做选择。如果 I 选择了升级，则核升级风险会进一步增加。如果还没有发生核战争，则轮到 II 做选择。以此轮替，直至其中一方做出让步或者发生核战争时博弈结束。对任何一方来说，获胜好过让步，让步好过核战争，而维持现状好过让步，即 $w>0>s>d$。

当一国选择坚持升级时，要么迫使对手先让步从而自己获胜，要么引发核战争。设引发核战争的总风险为 r，则一国选择升级的预期收益为 $w(1-r)+dr$。当升级的预期收益不低于让步的预期收益，即 $w(1-r)+dr\geq s$，亦即 $r\leq(w-s)/(w-d)$时，该国会选择升级。对于防御方 II 来说，它所愿意承受的引发核战争的最大风险 $R_{II}=(w_{II}-s_{II})/(w_{II}-d_{II})$ 即为它慑止进攻方进攻的决心。[③]

由上述模型可知：第一，危机中谁的决心更大，亦即谁能承受的核升级风险更高，谁就能在博弈中比对方坚持更多的轮次，从而迫使对方先让步；第二，在相互确保摧毁使得 d_I 和 d_{II} 值无显著差异的情况下，双方的决心对比只取决于 w 和 s 这两个参数，而这两个参数定义了双方各自在危机中的利害。总之，在相互确保摧毁

① Thomas C. Schelling, *Arms and Influence*, New Haven: Yale University Press, 1966, Chapter 3; Robert Powell, "The Theoretical Foundations of Strategic Nuclear Deterrence," *Political Science Quarterly*, Vol. 100, No. 1, 1985, p. 92; Robert Jervis, *The Meaning of the Nuclear Revolution*, Ithaca, NY: Cornell University Press, 1989; Robert Powell, *Nuclear Deterrence Theory: The Search for Credibility*, Cambridge: Cambridge University Press, 1990, pp. 33-109; Robert Powell, "Nuclear Brinkmanship, Limited War, and Military Power," *International Organization*, Vol. 69, No. 3, 2015, p. 594.

② 参见 Matthew Kroenig, "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes," *International Organization*, Vol. 67, No. 1, 2013, pp. 146-147。

③ Steven J. Brams, *Superpower Games: Applying Game Theory to Superpower Conflict*, New Haven: Yale University Press, 1985; Robert Powell, "Crisis Bargaining, Escalation, and MAD," *The American Political Science Review*, Vol. 81, No. 3, 1987, p. 721; Robert Powell, "Nuclear Deterrence Theory, Nuclear Proliferation, and National Missile Defense," *International Security*, Vol. 27, No. 4, 2003, p. 92; Robert Powell, "Nuclear Brinkmanship, Limited War, and Military Power," *International Organization*, Vol. 69, No. 3, 2015, p. 592.

条件下，核威慑能否成功取决于威慑方的相对决心大小，而相对决心大小又取决于威慑方在冲突中的相对利害大小。

四　核武器数量低于确保摧毁门槛时的核威慑原理

如果说“确保摧毁是有效核威慑的充分条件”还是相对少数派的观点，那么“前者是后者的必要条件”似乎不存在异议。但事实上，即使核武器数量低于确保摧毁门槛，同样有可能慑止敌国的常规和核进攻。确保摧毁不是有效核威慑的必要条件。

需要指出的是，现有文献中也有关于少量核武器即足以慑止核进攻的论述。[①] 如前文已经提及的，这些论述的核心逻辑主要立足于首次打击的不确定性，即认为在首次打击不能保证完全解除核报复能力的情况下，实现可靠核报复实际所需的核武器数量可以低于理论上的确保摧毁门槛。但即使是这种观点，似仍嫌未能充分把握核威慑生效的全部逻辑。事实上，核威慑生效所需的物质条件比这种观点所认为的还要宽松：即使不存在首次打击不确定性，即对手有百分之百的把握在首次打击中完全解除己方的核报复能力，并且这一点是双方的共同知识（common knowledge），己方同样有慑止敌方进攻的可能。

（一）威慑有效性≠战略稳定性

与威慑有效性相比，学界和战略界更经常谈论的概念是战略稳定性，并几乎毫无例外地将其与确保摧毁联系在一起。例如，北京大学中美人文交流研究基地和全美亚洲研究所协同十余家中美智库于 2016 年共同发布的《战略领域的中美关系》报告，将中美战略稳定性界定为“双方部署的核力量都能够在首次打击中生存下来，并且能够可信地向对方证明，彼此当前及未来的能力都无法阻绝对方的战略威慑”。[②] 撰写分报告的两位作者吴日强和埃尔布里奇·科尔比（Elbridge

① 参见 Bernard Brodie, *Strategy in the Missile Age*, Princeton, NJ: Princeton University Press, 1959, p. 275; Kenneth N. Waltz, “Toward Nuclear Peace,” in Robert J. Art and Kenneth N. Waltz, eds., *The Use of Force* (Second Edition), Lanham: University Press of America, 1983, pp. 584-598; John J. Mearsheimer, ‘‘Nuclear Weapons and Deterrence in Europe,’’ *International Security*, Vol. 9, No. 3, 1984/1985, pp. 19-46; Stephen J. Cimbala, “Deterrence Stability with Smaller Forces: Prospects and Problems,” *Journal of Peace Research*, Vol. 32, No. 1, 1995, pp. 65-78; Kenneth N. Waltz and Scott D. Sagan, *The Spread of Nuclear Weapons: A Debate*, New York: W.W. Norton, 1995, pp. 1-45。

② Elbridge A. Colby and Wu Riqiang, “Seeking Strategic Stability for U.S.- China Relations in the Nuclear Domain,” in Travis Tanner and Wang Dong, eds., *U.S.-China Relations in Strategic Domains*, The National Bureau of Asian Research, NBR Special Report #57, April 2016, p. 33.

Colby）认为，实现中美战略稳定的一个重要的必要条件是双方部署的核力量均可以在第一次打击中生存下来。[①] 邹治波和刘玮同样认为，维护中美核战略稳定的根本在于保持中国对美具有可信、可靠的第二次核打击能力。[②]

这里的问题是，相互确保摧毁是实现战略稳定的必要条件，但不是实现有效威慑的必要条件，战略稳定性不等于威慑有效性。战略稳定性是核威慑领域的一个专有概念，在本文所讨论的即时威慑情境下，特指危机稳定性，即双方常规军事冲突升级为核冲突的风险大小。[③] 威慑有效性是本文的一个工作概念，指防御方的威慑努力能否以及能在多大程度上慑止进攻方的进攻。乍一看，这两者似乎没有本质区别，但实际上，战略稳定不意味着威慑有效或者威慑成功，战略不稳定也不意味着威慑失效或者威慑不成功；将“战略稳定性”混同于“威慑有效性”，是导致许多学者误将相互确保摧毁视为核威慑有效的必要条件的一个可能原因。

如果将“稳定”理解为“不发生军事冲突”，那么广义的“稳定”显然包括“不发生常规军事冲突”（即“常规军事稳定”）和“不发生核冲突”（即“核稳定/战略稳定”）两个层面。然而，根据“稳定/不稳定悖论”（stability/instability paradox），这两个层面的“稳定”难以同时实现：由于核战争的成本不可承受，因此如果常规战争升级为核战争的风险高，那么冲突各方会注意管控冲突，从而使常规战争风险降低；如果常规战争升级为核战争的风险低，这意味着常规战争的预期成本低，由此又会使得核国家敢于更频繁地发动和介入常规战争。[④] 不妨设 ε 为常规战争升级为核战争的概率（亦即条件概率 p(常规战争升级为核战争|常规战争)），κ 为常规战

① 吴日强、[美] 埃尔布里奇·科尔比：《构建中美核武器领域战略稳定》，载《现代国际关系》2016 年第 10 期，第 50 页。

② 邹治波、刘玮：《构建中美核战略稳定性框架：非对称性战略平衡的视角》，载《国际安全研究》2019 年第 1 期，第 51 页。

③ Colin S. Gray, “Strategic Stability Reconsidered,” *Daedalus*, Vol. 109, No. 4, 1980, p. 135; Robert Jervis, *The Meaning of the Nuclear Revolution: Statecraft and the Prospect of Armageddon*, Ithaca, NY: Cornell University Press, 1989, pp. 138-140; Robert Powell, “Crisis Stability in the Nuclear Age,” *American Political Science Review*, Vol. 83, No. 1, 1989, pp. 61-76.

④ Glenn H. Snyder, “The Balance of Power and the Balance of Terror,” in Paul Seabury, ed., *The Balance of Power*, San Francisco, Calif.: Chandler, 1965, pp. 198-199; Robert Jervis, *The Illogic of American Nuclear Strategy*, Ithaca: Cornell University Press, 1984, p. 31; Robert Jervis, *The Meaning of the Nuclear Revolution*, Ithaca, N.Y.: Cornell University Press, 1989, p. 20; Charles L. Glaser, *Analyzing Strategic Nuclear Policy*, Princeton, N.J.: Princeton University Press, 1990, p. 46; Scott D. Sagan and Kenneth Waltz, *The Spread of Nuclear Weapons: A Debate Renewed*, New York: W. W. Norton, 1995; S. Paul Kapur, “India and Pakistan’s Unstable Peace: Why Nuclear South Asia is Not Like Cold War Europe,” *International Security*, Vol. 30, No. 2, 2005, p. 127.

争爆发的概率，那么核战争的风险 r＝κ×ε，对此“稳定/不稳定悖论”指出，ε 值的升高/降低会导致 κ 值的降低/升高。[①] 总之，核层面的稳定会引发常规层面的不稳定，核层面的不稳定会促成常规层面的稳定。[②]

现在假设有 A 和 B 两个有核国家，其中 A 国的核武器数量低于确保摧毁门槛，并且 A、B 两国都知道这一点。当 A、B 两国发生危机时，B 国显然有发动预防性打击预先消灭 A 国核武器的动机。A 国当然也能预见 B 国的这一动机，为避免自己有限的核武器在 B 国的首次打击下被全部消灭或者所剩无几以至于无法对 B 国造成显著伤害，A 国有动机抢先对 B 国发动先发制人式核打击。[③] 可见，一国核武器数量低于确保摧毁门槛的确会导致战略（核）层面的不稳定，亦即导致常规冲突升级为核冲突的风险（ε）升高。但根据“稳定/不稳定悖论”，ε 值升高将导致 κ 值降低，核战争风险（r）因此并不必然增加。与之对比，如果 A、B 两国核武器

① Robert Powell, “Nuclear Brinkmanship, Limited War, and Military Power,” *International Organization*, Vol. 69, No. 3, 2015, p. 596.

② 对“稳定/不稳定悖论”的形式化证明参见 Robert Powell, “Nuclear Brinkmanship, Limited War, and Military Power,” *International Organization*, Vol. 69, No. 3, 2015, pp. 589-626。统计证明参见 Daniel S. Geller, “Nuclear Weapons, Deterrence, and Crisis Escalation,” *The Journal of Conflict Resolution*, Vol. 34, No. 2, 1990, pp. 291-310; Robert Rauchhaus, “Evaluating the Nuclear Peace Hypothesis: A Quantitative Approach,” *The Journal of Conflict Resolution*, Vol. 53, No. 2, 2009, pp. 258-277; Matthew Kroenig and Robert F. Trager, “Nuclear Stability and Conventional Conflict,” manuscript, October 28, 2015, http://www.roberttrager.com/Research_files/KT2012.pdf; Bryan Early and Victor Asal, “Nuclear Weapons, Existential Threats, and the Stability–Instability Paradox,” *The Nonproliferation Review*, Vol. 25, No. 3/4, 2018, pp. 223-247。关于中美、印巴等有核国家安全关系进程符合“稳定/不稳定悖论”的论述，参见 Joshua Rovner, “Two Kinds of Catastrophe: Nuclear Escalation and Protracted War in Asia,” *Journal of Strategic Studies*, Vol. 40, No. 5, 2017, pp. 696-730; S. Paul Kapur, “India and Pakistan’s Unstable Peace: Why Nuclear South Asia is Not Like Cold War Europe,” *International Security*, Vol. 30, No. 2, 2005, pp. 127-152; Dinshaw Mistry, “Complexity of Deterrence among New Nuclear States: The India-Pakistan Case,” in T. V. Paul, Patrick M. Morgan and James J. Wirtz, eds., *Complex Deterrence: Strategy in the Global Age*, Chicago: The University of Chicago Press, 2009, pp. 183-203; Evan Braden Montgomery and Eric S. Edelman, “Rethinking Stability in South Asia: India, Pakistan, and the Competition for Escalation Dominance,” *Journal of Strategic Studies*, Vol. 38, No. 1-2, 2015, pp. 159-182。中国战略界对中美危机稳定性的理解也与“稳定/不稳定悖论”的逻辑相吻合，参见 Fiona S. Cunningham and M. Taylor Fravel, “Dangerous Confidence? Chinese Views on Nuclear Escalation,” *International Security*, Vol. 44, No. 2, 2019, pp. 61-109。当然，也有少数学者从“稳定”的界定和衡量标准角度对“稳定/不稳定悖论”提出质疑，参见胡高辰、李彬：《稳定—不稳定悖论的批判与美国的安全研究范式分析》，载《国际论坛》2018 年第 4 期，第 53-54 页。

③ Charles L. Glaser, *Analyzing Strategic Nuclear Policy*, Princeton, N.J.: Princeton University Press, 1990, p. 45.

数量均达到确保摧毁门槛，此时 ε 值固然会降低，但也会由此导致 κ 值升高。后一种情形下的 r 值因此并不必然低于前一种情形下的 r 值。

总之，与核武器数量达到确保摧毁门槛相比，低于确保摧毁门槛的确会增加战略不稳定性。但凭借由这种不稳定性所带来的稳定性，未达到确保摧毁门槛的核武器依然有可能将核战争的风险压制在与达到确保摧毁门槛时相当甚至更低的水平上。

（二）核武器数量低于确保摧毁门槛时的威慑承诺更可信

由于“稳定/不稳定悖论”，核武器数量即使低于确保摧毁门槛也仍然有慑止敌国进攻的能力，但这只是保证了核威慑生效的硬实力条件。如第三部分所述，核威慑要想生效，同样甚至更加重要的是核威慑承诺的可信度。第三部分已经指出，核武器数量与核威慑承诺可信度之间存在张力，当核武器数量达到确保摧毁门槛时，核威慑承诺会因此而在一定程度上变得不可信。由于以下两种机制，核武器数量与核威慑承诺可信度之间的这种张力反过来同样存在：核武器数量低于确保摧毁门槛会增加核威慑承诺的可信度。

首先，核武器数量低于确保摧毁门槛会压缩威慑方实施即时威慑的行动空间。根据现有理论，在完全谈判（pure bargaining）——如果谈判达成协议，谈判双方从协议中获得的效用大小将存在差异，但对双方来说，达成协议都好过达不成协议——条件下，如果其中一方的选择自由受到某种不可撤销的限制，这将反而使其在谈判中占据优势，增加按照对其更有利的方式达成协议的可能性。[①] 核危机状态下的进攻—威慑互动就是一种典型的完全谈判，而核武器低于确保摧毁门槛这一点恰恰剥夺了威慑方的一个重要策略选项：等待进攻方发动首次核打击后对其实施核报复。从信号传递的角度看，这种缩小行动空间的约束条件将使得威慑方使用核武器的威胁更为可信。[②]

在危机中，威慑方在得到进攻方即将发动核进攻或大规模常规进攻的警报或预判后，必须在两种各有利弊的选项中做出抉择：要么果断实施先发制人式打击，这样做将冒误判而爆发不必要核战争的风险，要么选择等待进攻情报核实后再实施报复，这样做将冒己方报复能力被对方首次打击摧毁的风险。[③] 显然，在拥有确保摧毁能力的

① Thomas C. Schelling, *The Strategy of Conflict*, Cambridge, MA: Harvard University Press, 1960, Chapter 2.

② 关于压缩行动空间有助于增加承诺可信度的经典论述，参见 James D. Fearon, “Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs,” *The Journal of Conflict Resolution*, Vol. 41, No. 1, 1997, pp. 68-90。

③ Thomas C. Schelling, *Arms and Influence*, New Haven: Yale University Press, 1966, pp. 228-229.

情况下，威慑方更敢于承担后一种风险。可靠的二次核打击能力固然给予了威慑方在危急时刻保持战略定力的底气，但同时也削弱了威慑方发动先发制人式核打击的决心。相反，那些核武器难以在首次打击中生存下来的国家无疑更愿意承担前一种风险，一旦认为有核战争风险，它们会有强烈的动机首先使用核武器，[①] 因为它们等不起。这类似两辆在狭窄道路上相向而行的汽车，如果其中一辆车的外侧是悬崖，而另一辆车的外侧是应急车道，那么几乎一定是后者给前者让路，因为后者知道，前者让路等于摔死，行动空间的劣势使前者比后者更愿意冒两车相撞的风险。

需要说明的是，在对方拥有确保摧毁能力的情况下，与核武器数量达到确保摧毁门槛时一样，未达到确保摧毁门槛时使用核武器同样是“自杀”行为。换言之，后一种情形下的核威慑承诺同样也不是百分之百可置信的。所谓“核武器数量低于确保摧毁门槛时的威慑承诺更可信”，是相对于达到确保摧毁门槛时的承诺可信度而言的。简而言之，两种情形下使用核武器的可信度都不高，但由于未达到确保摧毁门槛时威慑方的行动空间更小，这使得这种情形下威慑方冒险使用核武器的可能性——亦即威慑承诺可信度——相对更大。[②]

其次，核武器数量低于确保摧毁门槛会迫使威慑方尽快将实力较量转换为决心较量。如第三部分所述，在继续冲突将导致双方同归于尽的情况下，决定冲突胜负的关键因素是双方的决心。因此，实力较弱但比对手更有决心的威慑方，有动机主动使风险升级，将双方推向核战争的边缘，从而将双方的博弈从军事力量的较量（contest of military strength）拖入决心的较量（test of resolve）。[③] 对于那些核武器难以在首次打击中生存下来的国家来说，它们只有在自己首先使用核武器的情况下才有可能造成两败俱伤的局面，为了保证这个跟对手拼决心的“筹码”能够发挥作用，这类有核国家抢先升级、将冲突拖入决心较量的意愿会更急迫。

如前所述，危机中谁的利害更大，谁的决心就更大。一般而言，在有美国参与的地区性争端中，美国的利益通常“并不真正至关重要”，因此决心对比往往更有利于地区有核国家。[④] 这意味着，对于那些核武器数量未达到确保摧毁门槛，而与美国的冲

① Fiona S. Cunningham and M. Taylor Fravel, “Dangerous Confidence? Chinese Views on Nuclear Escalation,” *International Security*, Vol. 44, No. 2, 2019, pp. 70-71.

② 感谢《国际安全研究》匿名评审专家提醒笔者更清楚地阐明这一点。

③ Robert Powell, “Nuclear Brinkmanship, Limited War, and Military Power,” *International Organization*, Vol. 69, No. 3, 2015, pp. 593, 613.

④ Charles L. Glaser and Steve Fetter, “National Missile Defense and the Future of US Nuclear Weapons Policy,” *International Security*, Vol. 26, No. 1, 2001, p. 69.

突又关系民族或政权存亡的国家来说，抢在美国发动预防性打击前率先将冲突升级至发动先发制人式核打击的边缘，既能避自己核武器数量有限之短，又能扬自己决心更大之长。美国在预见到决心竞争将来得非常快并且自己大概率会输掉决心竞争这一点之后，其理性选择就只能是尽量避免挑起争端，从一开始就不给对手制造危机升级的机会。而这正是这些国家想要实现的威慑目标。正因如此，就连反复断言美国核优势、否认包括俄罗斯在内的其他有核国家核威慑有效性的利伯和普雷斯也不得不承认，在对手能够在首次打击中全部摧毁己方有效核打击力量的情况下，通过强制性核升级制造核僵局是保全自己、对抗核优势方的极少数有效手段之一。①

（三）经验证据

从上述分析可见，核武器数量低于确保摧毁门槛时的核威慑能力并不必然低于达到该门槛时的核威慑能力，而低于该门槛时的威慑承诺反而比达到该门槛时的承诺更可信。根据这两点，核武器数量低于确保摧毁门槛时的威慑有效性至少不会比达到该门槛时的威慑有效性差。这一理论推断有经验证据的支持。

纳兰将地区性有核国家的核姿态（nuclear posture）划分为催化（catalytic）、确保报复和不对称升级（asymmetric escalation）三类。② 其中，确保报复要求核武器必须有可生存的二次打击能力，而催化和不对称升级不必满足这一条件。纳兰用多项 Logit 模型，对 1861—2001 年具有政治相关性的二元数据集进行统计分析，结果发现，不对称升级策略是唯一能够显著慑止从常规军事冲突到核战争的各个武装强度冲突的策略。采取这种策略能够使国家在战争和次战争级别上遭到攻击的风险平均降低三到四倍。而确保报复策略与威慑效果之间不存在统计显著性。③ 这项研究从实证层面印证，不具备二次打击能力的核武器同样甚至可能更加具有威慑有效性。

“不对称升级”是指在对手尚仅有发动常规进攻迹象时，威慑方就威胁以核打击而不是常规打击对该进攻进行报复，通过将冲突迅速升级至使用核武器的程度，

---

① 参见 Keir A. Lieber and Daryl G. Press, “The New Era of Nuclear Weapons, Deterrence, and Conflict,” *Strategic Studies Quarterly*, Vol. 7, No. 1, 2013, pp. 5-7。

② 核姿态指一国核力量结构的总体设定，包括核弹头和运载工具的数量和类型、有关核武器何时、如何部署和使用以及针对什么目标的一套规则和程序等，可以近似地理解为核策略的操作化。参见 Vipin Narang, “What Does It Take to Deter? Regional Power Nuclear Postures and International Conflict,” *The Journal of Conflict Resolution*, Vol. 57, No. 3, 2012, p. 483。

③ Vipin Narang, “What Does It Take to Deter? Regional Power Nuclear Postures and International Conflict,” *The Journal of Conflict Resolution*, Vol. 57, No. 3, 2012, pp. 484-501; Vipin Narang, *Nuclear Strategy in the Modern Era: Regional Powers and International Conflict*, Princeton: Princeton University Press, 2014, Chapter 9.

慑止对手的常规和核进攻。为了实现这种不对称升级，威慑方会明确降低使用核武器的门槛，公开发出首先使用核武器的威胁，同时还可能将核武器的使用权限下放至较低层级，增加核武器因中下级军官的非理性冲动而在低水平常规军事冲突中被使用的概率。[①] 这种核姿态的本质是将核升级的风险点尽量提前，如上所述，这样做能够抢在对手发动针对军事力量的核打击之前，将双方的博弈转变为决心竞赛，这是这种核姿态能够显著发挥威慑功能的内在原理。

除了统计分析，一些不具备确保摧毁能力的国家在危机状态下成功慑止进攻的案例，同样证明了确保摧毁不是即时核威慑生效的必要条件。在 1998 年核试验之后的相当长时间里，印度在核武器的数量和质量上都明显优于巴基斯坦。在各种投送系统方面，巴基斯坦相比较于印度也有明显差距。更为重要的是，印度的地理纵深远大于巴基斯坦，印度可以锁定巴基斯坦的任意地点，而印度的很大一部分领土是巴基斯坦核武库所无法触及的。此外，印度的太空侦察能力和远程空中侦察能力同样远优于巴基斯坦。[②] 由于巴基斯坦的核力量不足以支持其实施确保报复策略，因此自 1998 年至今，其全部核武库都是按首先使用、并且在首次使用时就升级到战略级别进行配置，其核姿态呈现出典型的“不对称升级”特征。[③]

巴基斯坦核威慑的首要目标是慑止印度的常规军事进攻。[④] 为实现这一目标，巴基斯坦明确保留对印度首先使用核武器的策略选项，[⑤] 同时刻意降低使用核武器

① Vipin Narang, “What Does It Take to Deter? Regional Power Nuclear Postures and International Conflict,” *The Journal of Conflict Resolution*, Vol. 57, No. 3, 2012, pp. 486-491.

② Christoph Bluth, “India and Pakistan: A Case of Asymmetric Nuclear Deterrence,” *Korean Journal of Defense Analysis*, Vol. 22, No. 3, 2010, p. 398.

③ Christoph Bluth, “India and Pakistan: A Case of Asymmetric Nuclear Deterrence,” *Korean Journal of Defense Analysis*, Vol. 22, No. 3, 2010, p. 397; Vipin Narang, ‘‘Posturing for Peace? Pakistan’s Nuclear Postures and South Asian Stability,’’ *International Security*, Vol. 34, No. 3, 2010, pp. 38-78. 关于巴基斯坦未来将继续依赖不对称升级策略的讨论，参见 Diana Wueger, “Pakistan’s Nuclear Future: Continued Dependence on Asymmetric Escalation,” *The Nonproliferation Review*, Vol. 26, No. 5-6, 2019, pp. 449-463。

④ Major Gen. Muhammad Ali Durrani, (Rtd.), ‘‘Pakistan Strategic Thinking and the Role of Nuclear Weapons,’’ Cooperative Monitoring Center Occasional Paper 37, Sandia National Laboratories, July 2004, http://www.cmc.sandia.gov/cmc-papers/sand20043375p.pdf, quoted from Christoph Bluth, “India and Pakistan: A Case of Asymmetric Nuclear Deterrence,” *Korean Journal of Defense Analysis*, Vol. 22, No. 3, 2010, p. 398.

⑤ Lt. Gen. F. S. Lodhi (Rtd.), ‘‘Pakistan’s Nuclear Doctrine,’’ *Defence Journal*, April 1999, http://www.defencejournal.com/apr99/pak-nuclear-doctrine.htm, quoted from Christoph Bluth, “India and Pakistan: A Case of Asymmetric Nuclear Deterrence,” *Korean Journal of Defense Analysis*, Vol. 22, No. 3, 2010, p. 397.

的门槛。1999 年，巴基斯坦外交秘书沙姆沙德·艾哈迈德（Shamshad Ahmad）宣称：“我们将毫不犹豫地使用我们武器库中的任何武器来捍卫我们的领土完整。”[①] 2001 年，时任巴基斯坦战略规划部主任哈立德·基德维（Khalid Kidwai）明确提出了巴基斯坦使用核武器的四种情况，包括遭受领土入侵、军队遭受重创、经济遭到扼杀和政治因境外势力煽动而出现动荡。[②] 前两种包括常规军事威胁，而后两种则甚至属于非军事压力。与此同时，巴基斯坦还有意识地将危机状态下的核武器使用权限下放，以使其核姿态更具进攻性。[③]

这种降低核升级门槛、推高核失控风险的威慑策略，直接迫使印度放弃了原有的“孤注一掷”式的大规模报复威慑策略，转而实施“冷启动”（Cold Start）策略，即采取有限行动，将常规军事打击限制在巴基斯坦边境附近，避免给巴基斯坦提供核升级的借口和刺激。[④] 从宏观结果看，这种不对称升级姿态成功迫使印度在长达 20 年期间（1999—2019 年）的绝大部分时间里，对巴基斯坦保持了总体上的战略忍耐。[⑤] 导致这种战略忍耐的唯一原因是印度对巴基斯坦首先使用核武器的担忧。[⑥]

如果说进入 21 世纪后，巴基斯坦核武器数量有可能逐渐达到确保摧毁门槛，但在 1999 年卡吉尔战争之前，巴基斯坦只有几十枚低当量（5kt—10kt）的导弹匹

① Celia Dugger, “Atmosphere Is Tense as India and Pakistan Agree to Talks,” *New York Times*, June 1, 1999, quoted from Vipin Narang, *Nuclear Strategy in the Modern Era: Regional Powers and International Conflict*, Princeton: Princeton University Press, 2014, p. 270.

② Christoph Bluth, “India and Pakistan: A Case of Asymmetric Nuclear Deterrence,” *Korean Journal of Defense Analysis*, Vol. 22, No. 3, 2010, p. 397. 参见 David J. Karl, “Pakistan’s Evolving Nuclear Weapon Posture,” *The Nonproliferation Review*, Vol. 21, No. 3-4, 2014, p. 318; Evan Braden Montgomery and Eric S. Edelman, “Rethinking Stability in South Asia: India, Pakistan, and the Competition for Escalation Dominance,” *Journal of Strategic Studies*, Vol. 38, No. 1-2, 2015, p. 170。

③ David J. Karl, “Pakistan’s Evolving Nuclear Weapon Posture,” *The Nonproliferation Review*, Vol. 21, No. 3-4, 2014, p. 325; Evan Braden Montgomery and Eric S. Edelman, “Rethinking Stability in South Asia: India, Pakistan, and the Competition for Escalation Dominance,” *Journal of Strategic Studies*, Vol. 38, No. 1-2, 2015, p. 163.

④ Robert Powell, “Nuclear Brinkmanship, Limited War, and Military Power,” *International Organization*, Vol. 69, No. 3, 2015, pp. 615-616.

⑤ Devin T. Hagerty, *Nuclear Weapons and Deterrence Stability in South Asia*, Cham, Switzerland: Springer International Publishing, 2020, Chapters 2, 3.

⑥ V. K. Sood and Pravin Sawhney, *Operation Parakram: The War Unfinished*, New Delhi: Sage, 2003, p. 116, quoted from Robert Powell, “Nuclear Brinkmanship, Limited War, and Military Power,” *International Organization*, Vol. 69, No. 3, 2015, p. 616.

配弹头及少量短程和中程弹道导弹，[①] 几乎肯定不具备可靠的对印二次核摧毁能力。但在卡吉尔战争中，巴基斯坦的核武器依然成功慑止了印度对其边境渗透行动的大规模报复。1999 年 5 月，巴基斯坦“渗透者”进入印度境内并与印军交火。印度总理瓦杰帕伊及印度人民党领导层非常担心一旦印度军队越界进入巴基斯坦领土，巴基斯坦将首先使用核武器，因此不得不对印军作战策略作出限制，避免在实控线或双方边界开辟新的战线，并严令印军不得越过实控线在巴基斯坦境内采取任何行动。[②]

作为最新拥有核武器的国家，朝鲜核武库规模同样很难达到严格意义上的确保摧毁标准。根据美国国防部、韩国政府、瑞典斯德哥尔摩国际和平研究所、《原子科学家公报》等多方评估结果，2018 年朝鲜核武器数量约为 20—60 枚。[③] 正是由于核武器数量非常有限，朝鲜多年来始终坚持在常规军事危机中首先使用核武器以慑止军事入侵的核战略。[④] 2013 年，朝鲜制定核作战（nuclear warfighting）计划，主要内容包括使用弹道导弹攻击美国盟国及美国在东北亚的军事基地以慑止最初的攻击。[⑤] 当年 3 月，金正恩审议并批准了朝鲜人民军战略火箭部队可“随时打击美国本土及其在夏威夷、关岛和韩国的军事基地”的计划。[⑥] 随后，朝鲜以立法形式规定其核武器将“用于威慑和击退敌人对朝鲜的侵略和进攻”，[⑦] 同时表示其“核

① Vipin Narang, ‘‘Posturing for Peace?: Pakistan’s Nuclear Postures and South Asian Stability,’’ *International Security*, Vol. 34, No. 3, 2010, p. 57.

② Vipin Narang, *Nuclear Strategy in the Modern Era: Regional Powers and International Conflict*, Princeton: Princeton University Press, 2014, pp. 269, 271-272.

③ 谢瑞强、刘一鸣：《美估计朝鲜或拥有 60 枚核弹，半岛无核化走向何方？》，澎湃新闻，2020 年 8 月 22 日，https://m.thepaper.cn/newsDetail_forward_8831526；Hans M. Kristensen and Robert S. Norris, “North Korean Nuclear Capabilities, 2018,” *Bulletin of the Atomic Scientists*, Vol. 74, No. 1, 2018, https://thebulletin.org/2018/01/north-korean-nuclear-capabilities-2018。

④ Terence Roehrig, “North Korea’s Nuclear Weapons Program: Motivations, Strategy, and Doctrine,” in Toshi Yoshihara and James R. Holmes, eds., *Strategy in the Second Nuclear Age: Power, Ambition, and the Ultimate Weapon*, Washington, D.C.: Georgetown University Press, 2012, pp. 81-98.

⑤ Nick Kodama, “Threatening the Unthinkable: Strategic Stability and the Credibility of North Korea’s Nuclear Threats,” *Journal of Global Security Studies*, 2020, doi: 10.1093/jogss/ogaa004, p. 5.

⑥ Korean Central News Agency, “Nuclear War to Be Conducted on Korean Peninsula.” KCNA, March 29, 2013, https://kcnawatch.co/newstream/1451900772-347226436/nuclear-war-to-be-conducted-on-korean-peninsula/.

⑦ Korean Central News Agency, “Law on Consolidating Position of Nuclear Weapons State Adopted,” April 1, 2013, https://kcnawatch.co/newstream/1451896124-739013370/law-on-consolidating-position-of-nuclear-weapons-state-adopted/.

武装部队应始终保持战斗态势”。[①] 2016 年 3 月，联合国安理会通过一项新的对朝制裁决议，金正恩对此表示，“现在是我们把对敌人的军事反击模式转变为从各个方面进行先发制人的攻击的时候了”。[②]

正如分析人士普遍指出的，在常规军事力量和核力量与对手差距悬殊的情况下，将核门槛调低是朝鲜的理性选择。由于朝鲜核武库规模和运载工具数量都非常有限，因此当其感知到美国有可能发起进攻时，朝鲜有强烈动机发动先发制人式的第一次核打击。美国也非常清楚，一旦其对朝鲜的威胁稍稍过度，其在关岛和日韩等地部署的军事基地将有极大可能遭受朝鲜的突然核打击。因此对美国来说，维持朝鲜半岛稳定和自身安全的理性选择，就是尽量避免发生可能诱使朝鲜首先使用核武器的危机，从一开始就不给朝鲜升级冲突的理由。[③]

朝鲜核威慑的有效性在 2017 年经受住了迄今最严峻的考验。特朗普 2017 年上台后对朝采取了“极限施压”（Maximum Pressure）政策，[④] 该政策的一个重要特征是明确将进攻性军事手段纳入可选策略范畴。[⑤] 2017 年 3 月，美韩举行“关键决心”联合军事演习，参演兵力创美韩举行联合军演以来规模之最。[⑥] 4 月 28 日，特朗普表示，

① Shane Smith, “North Korea’s Evolving Nuclear Strategy,” research report, US-Korea Institute at SAIS, August 2015, p. 17, https://wmdcenter.ndu.edu/Portals/97/Documents/Publications/Articles/Evolving-Nuclear-Strategy.pdf.

② Korean Central News Agency, “Kim Jong Un Guides Test-fire of New Multiple Launch Rocket System,” March 4, 2016, https://kcnawatch.org/newstream/252244/kim-jong-un-guides-test-fire-of-new-multiple-launch-rocket-system/.

③ Nick Kodama, “Threatening the Unthinkable: Strategic Stability and the Credibility of North Korea’s Nuclear Threats,” *Journal of Global Security Studies*, 2020, doi: 10.1093/jogss/ogaa004, pp. 1-15; Jon Wolfsthal, “No Joke: When Donald Trump Hurls Insults, North Korea Thinks about War,” *USA Today*, September 26, 2017, https://www.usatoday.com/story/opinion/2017/09/26/donald-trump-hurls-insults-threats-north-korea-ponders-nuclear-war-jon-b-wolfsthal-column/701494001/; Vipin Narang, “Why Kim Jong Un Wouldn’t Be Irrational to Use a Nuclear Bomb First,” *Washington Post*, September 8, 2017, https://www.washingtonpost.com/outlook/why-kim-jong-un-wouldnt-be-irrational-to-use-a-nuclear-bomb-first/2017/09/08/a9d36ca4-934f-11e7-aace-04b862b2b3f3_story.html?utm_term=.40ff5c58bbef; Adam Mount and Mira Rapp-Hooper, “Nuclear Stability on the Korean Peninsula,” *Survival*, Vol. 62, No. 1, 2020, pp. 40-41.

④ Matthew Pennington, “As Tensions Rise in North Korea, Trump Administration Calls for ‘Maximum Pressure’,” *Time.com*, April 14, 2017, https://www.cbsnews.com/news/as-north-korea-threatens-nuclear-test-pence-has-foreign-policy-gaps-to-fill/.

⑤ Carol E. Lee, “Trump Issues New Warning to North Korea,” *Wall Street Journal* (Online), April 11, 2017; Zeke J. Miller, “Trump Administration: Military Preparations ‘Underway’ for North Korea,” *Time.com*, April 26, 2017.

⑥ 朱克川：《朝鲜半岛：特朗普无法翻云覆雨》，载《军事文摘》2017 年第 7 期，第 8 页。

“我们绝对有可能最终与朝鲜发生一场重大冲突”。[①] 5 月 19 日，“卡尔·文森”号和“罗纳德·里根”号航母首次同时参加美韩联合演习。[②] 8 月 11 日，特朗普发推特称，美国“军事方案已准备就绪，装弹上膛，锁定目标。希望金正恩另寻明路”。[③]

面对美国的军事威胁，朝鲜采取了一系列故意推高战争风险的“挑衅”行动。自 2017 年 2 月至 11 月，朝鲜共试射各种短、中、远程和洲际弹道导弹、反舰巡航导弹 16 次 24 枚，并于 9 月 3 日进行第六次核试验。[④] 在双方剑拔弩张的对峙中，最终美国首先作出了让步。2017 年 11 月 7 日，特朗普在访问韩国期间明确表示，美国解决朝核问题的各种可选工具中不包括军事行动；[⑤] 11 月 28 日，美国国务卿蒂勒森甚至表示，美国“已经准备好无条件”与朝鲜领导人会面。[⑥] 这次朝核危机最终以和平且有利于朝鲜的方式结束。巴基斯坦和朝鲜这样以极少数核武器成功慑止强敌进攻的案例至少表明，不具备确保摧毁能力并不一定意味着无法实现有效核威慑。

五　不同核武器数量水平下的核威慑生效机制

由以上两部分分析可知：第一，仅靠核武器自身无法自动形成有效核威慑；[⑦] 第二，核武器数量是否达到确保摧毁门槛与核威慑是否有效无关；第三，核武器数量

① Gerry Mullany, “Trump Warns that ‘Major, Major Conflict’ with North Korea is Possible,” *New York Times*, April 28, 2017, https://www.nytimes.com/2017/04/27/world/asia/trump-north-korea-kim-jong-un.html?smprod=nytcore-iphone&smid=nytcore-iphone-share.

② 夏立平：《双层博弈理论视阈下特朗普政府的朝核政策》，载《美国研究》2017 年第 6 期，第 127 页；樊吉社：《特朗普政府对朝政策逻辑与朝核问题前景》，载《现代国际关系》2017 年第 7 期，第 19-20 页。

③ Paul D. Shinkman, “Trump: Military ‘Locked and Loaded’ to Counter North Korea,” *U.S. News & World Report*, August 11, 2017, https://www.usnews.com/news/national-news/articles/2017-08-11/trump-military-solutions-for-north-korea-locked-and-loaded.

④ “North Korean Missile Launches & Nuclear Tests: 1984-Present,” Center for Strategic and International Studies, https://missilethreat.csis.org/north-korea-missile-launches-1984-present/.

⑤ “North Korea: How to Approach the Nuclear Threat,” Seoul, November 7, 2017, https://www.whitehouse.gov/briefings-statements/remarks-president-trump-president-moon-republic-korea-joint-press-conference-seoul-republic-korea/.

⑥ George Perkovich, “Can Trump Cut a Deal With North Korea?” *Politico*, December 13, 2017, https://www.politico.com/magazine/story/2017/12/13/can-trump-cut-a-deal-with-north-korea-216072.

⑦ David G. Coleman and Joseph M. Siracusa, *Real-World Nuclear Deterrence: The Making of International Strategy*, Westport: Praeger Security International, 2006, p. 3; Keir A. Lieber and Daryl G. Press, *The Myth of the Nuclear Revolution: Power Politics in the Atomic Age*, Ithaca: Cornell University Press, 2020, p. 5.

达到和未达到确保摧毁门槛时的核威慑生效机制不同。在此基础上，本部分将对一国核武器数量的水平作出更细致的划分，以进一步厘清不同核武器数量水平下影响核威慑生效的关键因素和提高核威慑有效性的可选途径。

根据现有对核威慑的认识，一国核武器数量在增长过程中存在三个关键节点。如图 1 的数轴所示，在节点 a 处，一国核武器达到在己方首先使用且遭到对方拦截的情况下能够给对方造成不可承受的损失的最低数量；在节点 b 处，达到对方首次打击所能消灭以及对方拦截系统所能拦截的数量上限之和；在节点 c 处，达到在遭到对方首次打击后对对方实施报复能够给对方造成不可承受的损失的最低数量。

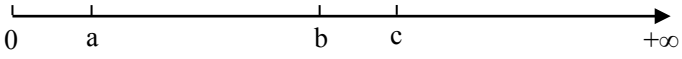


图 1　核武器数量阶段划分

资料来源：作者自制。

关于这三个节点，有两点需要说明：其一，在现实中，互动双方对其中一方核武器数量是否达到某个节点的主观判断可能存在差异，这种判断差异有可能影响核威慑的效果。但为便于将注意力聚焦于最核心的威慑原理，这里假定一国核武器数量是否达到某个节点是该国与其对手之间的共同知识，双方不存在认识分歧或信息不对称问题。

其二，在现实中，a 与 0 的差值以及 c 与 b 的差值不会太大。这里涉及对多少枚核武器才能造成不可承受的损失的判断。按照麦克纳马拉的标准，制造不可承受的损失需要 200—400 枚核武器。[①] 但正如许多学者所指出的，决策者不可承受损失的实际阈值远低于该标准。例如邦迪认为，对于真实世界的领导人来说，任何使本国遭受哪怕一枚核弹打击的决定都会从一开始就被认为是“灾难性的错误”。[②] 格拉泽等同样认为，在实际决策中，如果美国决策者认为某个行动会导致哪怕一个美国城市被摧毁，这个行动就几乎不可能被实施。[③] 柯庆生（Thomas J. Christensen）甚至认为，中国即使没有能够击中美国本土的洲际导弹，只要有能力打击美国在亚太地区的盟友以及美国在日韩等国部署的基地，就足以被美国领导层视为能够制造

① 参见 Alain C. Enthoven and K. Wayne Smith, *How Much Is Enough? Shaping the Defense Program, 1961-1969*, Santa Monica: RAND Corporation, 2005, p. 207。

② McGeorge Bundy, “To Cap the Volcano,” *Foreign Affairs*, Vol. 48, No. 1, 1969, p. 10.

③ Charles L. Glaser and Steve Fetter, “Should the United States Reject MAD? Damage Limitation and U.S. Nuclear Strategy toward China,” *International Security*, Vol. 41, No. 1, 2016, p. 59.

不可承受的损失。[①] 总之，从 0 到 a 和从 b 到 c 所需的增量不大。

上述三个节点将一国核武器数量水平划分为四个阶段，阶段名称及各阶段核威慑生效方式如表 1 所示。其中，区间[0, a)是无有效威慑阶段。在该阶段，由于无法给对手造成不可承受的损失，因此理论上不具备核威慑能力。但如上所述，这个区间的实际距离很短，有核国家不难实现从没有核威慑能力到有核威慑能力的跨越。

表 1　不同核武器数量水平下的核威慑生效方式

核武器数量水平	战略稳定性	核威慑生效途径	核心举措	威慑原理
无有效威慑阶段 [0, a)	低	—	—	—
战略不稳定阶段 [a, b]	低	不对称升级	核升级风险点提前	“稳定/不稳定悖论”
过渡阶段 (b, c)	中	不对称升级，如果对手采取核优势战略	核升级风险点提前	“稳定/不稳定悖论”
		核边缘，如果对手采取确保摧毁战略	制造不能完全掌控的风险	核边缘机制
战略稳定阶段 [c, +∞)	高	核边缘，如果突发大规模核战争风险高	建立核武器数量优势	核边缘机制
		确保对等报复，如果突发大规模核战争风险低		消耗战

表格来源：作者自制。

区间[a, b]是战略不稳定阶段。在该阶段，如前所述，由于核武器数量未达到确保摧毁门槛，对手有动机发动预防性针对军事力量的核打击，[②] 因此互动双方的战略稳定性（危机稳定性）低，但战略不稳定不等于威慑无效，核武器数量处于该阶段的有核国家同样有机会成功慑止对手的进攻。当然，要想增加威慑成功的几率，该阶段的有核国家必须通过“随时准备发动核打击”这样激进强硬的政策宣示和相应的军备姿态，将核升级的风险点尽可能提前。在这个阶段，“不首先使用核武器”这种单边自我限制的政策承诺不仅缺乏战略意义，而且非常危险。

区间(b, c)是过渡阶段。与前后两个阶段相比，该阶段的核威慑最为脆弱。一方

① Thomas J. Christensen, “The Meaning of the Nuclear Evolution: China’s Strategic Modernization and US-China Security Relations,” *Journal of Strategic Studies*, Vol. 35, No. 4, 2012, pp. 459-460.

② 当然，在区间[0, a)时对方同样也有这种动机，无有效威慑阶段可视为广义上的战略不稳定阶段。

面，该阶段核武器数量已经超出了对方首次打击所能消灭以及对方所能拦截的核武器数量上限，根据前文的分析，该阶段核威慑的意图可信度（承诺可信度）会因此低于此前的战略不稳定阶段；另一方面，该阶段核武器数量又尚未达到确保摧毁门槛，二次核打击无法给对手造成不可承受的损失，因此该阶段核威慑的能力可信度又低于其后的战略稳定阶段。总之，核武器数量处于该阶段的有核国家既没有硬实力优势，又没有承诺可信度优势，要想实现有效核威慑，需根据对手核战略的不同而选择不同的威慑策略。

如果对手采取的是核优势（nuclear primacy）或者损伤限制（damage limitation）这样旨在通过首先发动预防性核打击削弱甚至消灭对方核武器的进攻性核战略，[①] 那么过渡阶段有核国家面临的处境与其在战略不稳定阶段所面临的处境类似，即如果不能抢在对方使用核武器之前首先使用核武器，那么将无法给对方造成不可承受的损失，从而失去将实力竞争转换为决心竞争的机会。因此，这种情况下实现有效核威慑的机制与战略不稳定阶段相似，需依靠“不对称升级”这样将核升级风险点提前的进攻性核姿态，以“攻”对“攻”，迫使对手从一开始就避免挑起争端。

如果对手采取的是确保摧毁或确保报复这样旨在通过威胁报复性核打击以慑止对手进攻的防御性核战略，那么边缘策略是过渡阶段有核国家可以选择和依赖的最主要的即时威慑手段。如第三部分所述，边缘策略实施的关键是将局势拖入随时可能失控的不稳定状态。而要想比较容易地达到这种状态，需要互动双方至少有一方的核武器是脆弱的，这样才能使常规军事冲突滑向核冲突的前景变得可信。[②] 过渡阶段的核武器数量恰恰满足边缘策略对脆弱性的要求。在二次打击能力有限的情况下，危机中威慑方发动先发制人式核打击的意愿会比有充足二次打击能力时更强烈更可信。许多美国学者都指出，在中国拥有对美二次打击能力但军事实力处于弱

① 核优势战略强调首次打击能力（first-strike capability），其目标是在首次预防性核打击中将对手的所有核武器全部消灭。损伤限制战略相对放宽了对首次打击的要求，旨在在首次打击中尽可能削弱对手的核报复能力，使其剩余核能力不足以对己方造成不可承受的损失。参见 Tom Sauer, “A Second Nuclear Revolution: From Nuclear Primacy to Post-Existential Deterrence,” *Journal of Strategic Studies*, Vol. 32, No. 5, 2009, p. 746; Kier A. Lieber and Daryl G. Press, “The New Era of Nuclear Weapons, Deterrence, and Conflict,” *Strategic Studies Quarterly*, Vol. 7, No. 1, 2013, p. 8; Charles L. Glaser and Steve Fetter, “Should the United States Reject MAD? Damage Limitation and U.S. Nuclear Strategy toward China,” *International Security*, Vol. 41, No. 1, 2016, pp. 54-59。

② Robert Powell, “The Theoretical Foundations of Strategic Nuclear Deterrence,” *Political Science Quarterly*, Vol. 100, No. 1, 1985, pp. 77-80.

势的情况下，中国采取边缘策略将对美国构成强制性压力。[①] 被威慑方的这种忧患意识从另一个角度体现了边缘策略对威慑方的价值。

边缘策略与不对称升级策略的生效机制存在微妙差异：后者需明确威胁首先使用核武器，使被威慑方主要因惧怕首先遭受核打击而放弃进攻；前者在是否首先使用核武器这一点上相对模糊，更多依靠放大灾难性核交换的前景来慑止进攻，核交换中开“第一枪”的既有可能是威慑方也有可能是被威慑方。尽管存在这种区别，但这两种策略的生效机制有一个共同要点，即都必须故意增加核升级风险。这个要点深刻反映了“稳定/不稳定悖论”所揭示的“稳定与不稳定相互转化”原理。事实上，在中国战略界也有观点指出，“降低核阈值”是“核导弹部队军事威慑的主要手段”；[②] 在核强国依靠其在高技术常规武器方面的绝对优势进行一系列中高层空袭，而我方又无良好防范手段时，可适当降低核阈值，及时调整我核威慑政策，主动实施强有力的核威胁。[③] 这些观点正是上述核威慑原理的政策运用。

区间[c, +∞)是战略稳定阶段。在该阶段，核武器数量达到确保摧毁门槛，互动双方因此达到相互确保摧毁的战略稳定状态。既然如此，从直觉上看，这一阶段提高核威慑有效性的重点似乎就应当放在除增加核武器数量以外的其他方面，而不必再增加核武器。然而，综合现有相关理论和经验研究的发现可知，这一阶段核威慑的实际生效原理与这种直觉恰好相反：增加核武器数量是战略稳定阶段确保即时威慑可靠性的重要途径。

首先，在相互确保摧毁且大规模核战争爆发风险高的情况下，核武器数量对比能够显著影响危机双方的决心对比。如前所述，在相互确保摧毁状态下，可以通过边缘策略，将危机推向随时可能引发大规模核战争的失控边缘以实现可信威慑；在

① Thomas J. Christensen, “The Meaning of the Nuclear Evolution: China’s Strategic Modernization and US-China Security Relations,” *Journal of Strategic Studies*, Vol. 35, No. 4, 2012, pp. 460-463; Avery Goldstein, “First Things First: The Pressing Danger of Crisis Instability in U.S.-China Relations,” *International Security*, Vol. 37, No. 4, 2013, pp. 86-87; Caitlin Talmadge, “Would China Go Nuclear?: Assessing the Risk of Chinese Nuclear Escalation in a Conventional War with the United States,” *International Security*, Vol. 41, No. 4, 2017, pp. 50-92.

② 赵锡君：《慑战：导弹威慑纵横谈》，北京：国防大学出版社 2003 年版，第 34 页，转引自 Thomas J. Christensen, “The Meaning of the Nuclear Evolution: China’s Strategic Modernization and US-China Security Relations,” *Journal of Strategic Studies*, Vol. 35, No. 4, 2012, p. 479。

③ 于际训主编：《第二炮兵战役学》，北京：中国人民解放军出版社 2004 年版，第 294 页，转引自 Thomas J. Christensen, “The Meaning of the Nuclear Evolution: China’s Strategic Modernization and US-China Security Relations,” *Journal of Strategic Studies*, Vol. 35, No. 4, 2012, p. 477。

边缘状态下，决定威慑成败的关键是双方的决心对比。根据第三部分的模型，决心 R=(w-s)/(w-d)，由于相互确保摧毁，因此一般理论假定博弈双方在发生大规模核战争时的支付（d）无显著差异，在这种情况下两国决心（R）的相对大小由各自赢得危机的支付（w）和在危机中让步的支付（s）决定。但马修·克罗尼格（Matthew Kroenig）在这个一般理论的基础上进一步指出，即使相互确保摧毁，危机双方核武器的数量对比仍然可能影响双方在发生大规模核战争时各自的支付值（d_I 和 d_{II}），进而影响双方的决心对比。①

一方面，核武器更多的一方能够在核交换中对另一方造成更大程度和范围的摧毁；另一方面，核武器更多的一方能够在针对军事力量的打击中摧毁对方更多的核力量，从而减少自己遭受对方核打击时的损失。这两方面共同影响着双方对大规模核战争中彼此支付（d）的预期，对于预见到自己在可能的大规模核战争中损失更大的一方来说，平均而言将比另一方更愿意首先做出让步。这类似在狭窄道路上相向而行的两辆重量不等的汽车，尽管相撞对两辆车而言都是灾难，但较小较轻的一方先避让的可能性更大。克罗尼格的统计研究也表明，核武器数量更多的一方更有可能在危机中获胜，且核武器数量的相对优势越大，获胜的概率越高。②

其次，在相互确保摧毁且突然爆发大规模核战争风险低的情况下，核武器数量对比直接决定危机双方“核消耗战”的资源对比。如前所述，边缘策略生效的重要条件是至少有一方的核武器是相对脆弱的。由于这种脆弱性，脆弱一方首先使用核武器和遭受核打击后报复使用核武器对对方造成的损伤存在差异，这使得双方都有较强的先发制人动机，这种动机在很大程度上保证了大规模核战争的失控性风险，而只有当这种失控性风险存在且足够大时，边缘策略才能够发挥强制性效果。当核武器数量超过确保摧毁门槛并且进一步增加时，上述这种脆弱性以及由此导致的先发制人动机就都会减弱（因为首次打击和二次打击的效果的差异越来越小），在这种情况下，当面对可能的核打击时，双方将变得越来越有耐心，即使遭受小当量核

① Matthew Kroenig, “Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes,” *International Organization*, Vol. 67, No. 1, 2013, pp. 141-171; Matthew Kroenig, *The Logic of American Nuclear Strategy: Why Strategic Superiority Matters*, New York: Oxford University Press, 2018.

② Matthew Kroenig, “Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes,” pp. 148-162; Matthew Kroenig, *The Logic of American Nuclear Strategy: Why Strategic Superiority Matters*, New York: Oxford University Press, 2018, chapter 2, 3.

武器打击也有可能只会引发小当量核武器的反击，而不会立即引发大规模灾难性核交换，[①] 此时边缘策略就将变得不再适用。[②]

在上述这种危机难以迅速引发大规模核战争的情况下，核强制的主要机制就由此前的风险承受逻辑（risk-taking）转变为成本承受逻辑（cost-taking），双方的博弈在很大程度上变成了一场“消耗战”（war of attrition）：通过发动包括低当量战术核打击在内的一切非大规模核打击手段，向对方发出核打击有可能进一步升级的信号，[③] 对方如果拒绝退让，将同样选择发动对等性的战术核报复，由此形成循环，[④] 最终谁先让步取决于双方的核武库规模和对损失的承受能力。在损失承受能力相近时，双方可用的核武器数量对比就将成为决定冲突结果的最重要因素。[⑤] 迄今仅有的一个核武器用于实战以慑止（常规）进攻的案例就体现了这种“消耗战”逻辑：1945 年 8 月 6 日，广岛遭美国第一颗原子弹打击后，日本决策层一度猜测美国可能只有这一枚原子弹，直到 3 天后长崎再次遭受核打击，这种幻想才彻底破灭。假如美国当时真的只有一枚原子弹，那么日本投降的时间几乎必然会延迟。[⑥]

再次，确保摧毁的核武器数量门槛有可能随时间推移而提高。正如有学者所指出的，如果实现相互确保摧毁就能确保核威慑生效，就能确保双方生存安全，那么冷战后期美苏两国持续开展的大规模核军备竞赛就是令人困惑甚至是非理性的。[⑦]

---

① 冷战时期，美国战略界关于核升级是否可控（亦即有限核战争是否可行）的争论的一个核心要点就是，己方的核力量以及核指挥和控制系统能否在对方的首次有限核打击下生存下来。支持有限核战争的一方指出，双方核力量和指挥控制系统的生存可靠性越高，双方对对方可能的核打击就越有信心和耐性延缓做出反应，从而越有可能避免失控性核升级。对相关争论的梳理参见 Fiona S. Cunningham and M. Taylor Fravel, “Dangerous Confidence? Chinese Views on Nuclear Escalation,” *International Security*, Vol. 44, No. 2, 2019, pp. 66-72, especially at p. 68。

② Robert Powell, “The Theoretical Foundations of Strategic Nuclear Deterrence,” *Political Science Quarterly*, Vol. 100, No. 1, 1985, pp. 76-83.

③ Vince A. Manzo and John K. Warden, “After Nuclear First Use, What?” *Survival*, Vol. 60, No. 3, 2018, p. 138.

④ 这种核威慑策略被称为“确保对等报复”（assured reciprocal retaliation）。

⑤ Robert Powell, *Nuclear Deterrence Theory: The Search for Credibility*, Cambridge: Cambridge University Press, 1990, Chapter 7.

⑥ Herbert Feis, *The Atomic Bomb and the End of World War II*, Princeton: Princeton University Press, 1966, pp. 199-200; Wilson D. Miscamble C.S.C., *The Most Controversial Decision: Truman, the Atomic Bombs, and the Defeat of Japan*, New York: Cambridge University Press, 2011, pp. 97-98.

⑦ Brendan R. Green and Austin Long, “The MAD Who Wasn’t There: Soviet Reactions to the Late Cold War Nuclear Balance,” *Security Studies*, Vol. 26, No. 4, 2017, p. 607.

利伯和普雷斯在其最新专著中回答了这个困惑，其中一个重要结论是，相互确保摧毁形成的核僵局并非一劳永逸、不可逆转，对手核战略的变化以及新科技的运用，都有可能使一国原本足以确保摧毁的核能力变得不再够用。①

通过上述对不同核武器数量水平下核威慑生效机制的分析，我们可以对现有关于中国核武器数量的政策争论作出更深刻的反思。如第二部分所述，现有争论的一个基本共识是：如果核武器数量尚未达到确保摧毁门槛，那么应当增加核武器数量；如果已经达到，则无需增加。然而，以上分析的结论与这种共识恰好相反：如果核武器数量尚未达到确保摧毁门槛，那么维持这种数量水平同样甚至更加有可能实现有效核威慑；如果已经达到，则反而应当（继续）增加核武器。这个违背一般常识的结论是核武器数量与核威慑有效性之间关系的一个似非而是的悖论。

六 结论

《孙子兵法》有云：“兵者，国之大事，生死之地，存亡之道，不可不察也。”②在当今时代，核威慑大概是最需战略家深“察”细“察”的“国之大事”之一。但令人不安的是，当前战略界对这个关系国家生死存亡的“大事”仍然有所“不察”，这种“不察”集中体现在对“确保摧毁”原则的无批判接受：一方面认为在未达到确保摧毁门槛时努力增加（可生存的）核武器数量或者努力提高核武器在首次打击中的生存概率以达到确保摧毁门槛是实现有效核威慑的唯一途径，另一方面认为一旦达到确保摧毁门槛就可以放心地停止增加核武器。本文的讨论表明，这两种认识与核威慑的真实原理和经验事实不符；只有超越“确保摧毁”范式，才有可能更准确更完整地理解核威慑生效的全部机制。

本文的讨论提示，与核威慑相关的一些重要战略问题存在进一步研究的空间。一是如何在坚持“不首先使用核武器”的情况下实现有效核威慑。自 1964 年中国首枚原子弹试爆成功至今，中国政府始终坚持“不首先使用核武器”政策。③ 中国战略界的

① Keir A. Lieber and Daryl G. Press, *The Myth of the Nuclear Revolution: Power Politics in the Atomic Age*, Ithaca: Cornell University Press, 2020.

② 《孙子兵法 • 孙膑兵法》，骈宇骞等译注，北京：中华书局 2006 年版，第 3 页。

③ Jeffrey G. Lewis, “Chinese Nuclear Posture and Force Modernization,” *The Nonproliferation Review*, Vol. 16, No. 2, 2009, pp. 197-209; 樊吉社：《中国核政策的基本逻辑与前景》，载《外交评论》2018 年第 5 期，第 4-5 页。

主流观点对此持支持态度，[①] 但近年来也开始出现反思的声音。[②] 单纯从核威慑有效性的角度看，在多数情况下，“不首先使用核武器”这个承诺越可信，越不利于威慑的生效。[③] 如果中国选择坚持这个承诺，那么亟须研究在坚持这个承诺的前提下实现有效核威慑的原理和途径，这是一个在理论上存在很大空白因而难度极大的问题。二是如何确保确保报复战略的有效性。学界主流观点认为中国目前采取的核威慑战略是确保报复。[④] 但仅有确保报复的硬实力并不足以慑止进攻，且确保报复战略的实际威慑效果并不理想。如果中国坚持采取这个战略，需深入探索该战略实现有效核威慑的可能机制和所需条件。三是确定战略稳定性和威慑有效性两者的偏好排序。现有研究谈论战略稳定性的极多而专研威慑有效性的极少，而如本文所述，战略稳定性不仅不等于威慑有效性，而且在许多情况下两者存在张力，不可兼得。这提示我们在研究核战略时，须明确战略目标究竟是战略稳定性还是威慑有效性，如果将首要目标设定为战略稳定性，需考虑该战略对威慑有效性可能产生的负面影响。

【来稿日期：2021-03-25】
【修回日期：2021-05-06】
【责任编辑：李水生】

① 关于中国不首先使用核武器的论述和讨论，参见潘振强：《中国不首先使用核武器问题研究》，载李彬、赵通主编：《理解中国核思维》，北京，社会科学文献出版社 2016 年版，第 28-52 页；Christopher T. Yeaw, Andrew S. Erickson, and Michael S. Chase, “The Future of Chinese Nuclear Policy and Strategy,” in Toshi Yoshihara and James R. Holmes, eds., *Strategy in the Second Nuclear Age: Power, Ambition, and the Ultimate Weapon*, Washington, DC: Georgetown University Press, 2012, pp. 53-80。

② 相关梳理参见 Zhenqiang Pan, “A Study of China’s No-First-Use Policy on Nuclear Weapons,” *Journal for Peace and Nuclear Disarmament*, Vol. 1, No. 1, 2018, pp. 131-132。国际学术界在学理层面对“不首先使用核武器”政策也存在争论，支持的观点参见 David Gompert, Kenneth Watman and Dean Wilkening, “Nuclear First Use Revisited,” *Survival*, Vol. 37, No. 3, 1995, pp. 27-44; Steve Fetter and Jon Wolfsthal, “No First Use and Credible Deterrence,” *Journal for Peace and Nuclear Disarmament*, Vol. 1, No. 1, 2018, pp. 102-114；质疑的观点参见 Alexander Lanoszka and Thomas Leo Scherer, “Nuclear Ambiguity, No-First-Use, and Crisis Stability in Asymmetric Crises,” *The Nonproliferation Review*, Vol. 24, No. 3-4, 2018, DOI: 10.1080/10736700.2018.1430552; Brad Roberts, “Debating Nuclear No-First-Use, Again,” *Survival*, Vol. 61, No. 3, 2019, pp. 39-56。

③ 如表 1 所示，只有在核武器数量远远超过确保摧毁门槛、从而使得大规模核战争突发风险很低这一种情况下，核威慑有效性才不依赖于主动增加核升级风险（其前提是不承诺不首先使用核武器）这种途径。目前中国核武器数量可能尚未达到这个水平。

④ M. Taylor Fravel and Evan S. Medeiros, “China’s Search for Assured Retaliation: The Evolution of Chinese Nuclear Strategy and Force Structure,” *International Security*, Vol. 35, No. 2, 2010, pp. 48-87; Michael S. Chase, “China’s Transition to a More Credible Nuclear Deterrent: Implications and Challenges for the United States,” *Asia Policy*, No. 16, 2013, pp. 69-101; Fiona S. Cunningham and M. Taylor Fravel, “Assuring Assured Retaliation: China’s Nuclear Posture and U.S.-China Strategic Stability,” *International Security*, Vol. 40, No. 2, 2015, pp. 7-50.