

一体化与区域合作专题

# 国家间一体化： 主权让渡与红利预期的权衡

欧阳俊 邱 琼

**内容提要：**功能主义和政府间主义是国家间一体化理论的主流，两大学派之争贯穿着整个国际一体化实践与研究的历史，至今未休。虽然两大阵营都承认对方有合理的一面，自身存在某些局限甚至缺陷，各自都只是关于国家间一体化不同视角的描述和解释，但少有学者尝试将二者进行整合。本文的目的就在于整合上述两大理论，构建可解释不同区域不同类型国家间一体化的一般理论。本文首先回答了什么是国家间一体化，如何测度国家间一体化，明确地将其等同于国家主权让渡与汇集。其次，本文基于主权让渡和红利预期的权衡假设，借鉴国家建构理论的思路，构建了国家间一体化两阶段重复博弈模型，并在此基础上讨论了国家间主权汇聚与让渡的动力和条件，以及一体化演进/退化的反馈机理。最后，本文对全文进行总结，并指出本文构建的国家间一体化博弈模型的不足之处以及未来的研究方向。

**关键词：**国家间一体化 主权让渡 红利预期 两阶段博弈 演进/退化

**作者简介：**欧阳俊，西南财经大学经济与管理研究院，教授；邱琼，国家药品监督管理局，经济学博士。

**中图分类号：**D99 **文献标识码：**A

**文章编号：**1002-6649 (2020) 03-0001-26

斯坦利·霍夫曼于1966年发表《固执还是过时？民族国家的命运与西欧案例》一文后，功能主义与政府间主义之辩就成为国际一体化理论界的主旋律，一直延续到现在<sup>①</sup>。最初，政府间主义只是“王位的挑战者”。新功能主义尽管在“黑暗时代”遭受严重挫折，但直到20世纪90年代依旧是解释和指导欧洲一体化实践的主流理论。1998年，安德鲁·莫拉维斯克出版《欧洲的选择》，对传统政府间主义的假设进行了放松和修正，提出自由政府间主义理论，给予一体化参与者行为解释以微观理论基础，打破了原有的理论竞争格局<sup>②</sup>。此后20年，自由政府间主义获得广泛传播和接受，在与新功能主义竞争中隐隐领先。2019年莫拉维斯克曾自豪地宣称，“很少有学者或者评论家会质疑自由政府间主义对于欧洲一体化过去的解释能力。”即使竞争对手也不得不承认，“《欧洲的选择》是欧洲一体化研究领域被引用次数最多的文献之一”<sup>③</sup>。2008年以来，欧盟面临欧债危机、移民危机、英国脱欧、东西欧分歧公开化等诸多困境，一体化进程不仅陷入停滞，甚至有倒退的危险。在此背景下，2009年胡格等人强调了排他性身份认同与多层治理收益之间的张力，提出后功能主义一体化理论，成为与自由政府间主义抗衡的功能主义新辩手。

在半个多世纪的争辩过程中，也曾有学者尝试对此两大学派进行理论整合。譬如，托马斯·格林曾尝试在国际体制理论框架下，通过引入时间动态、反馈机制以及制度视角，从而实现基于自由政府间主义对于新功能主义的整合<sup>④</sup>。但是，功能主义者和政府间主义者彼此成见之深，在关于一体化特征和动态的辩论中总是严格按学派站队，理论整合工作并没有得到应有的关注和

<sup>①</sup> D. J. Puchala, “Institutionalism, Intergovernmentalism and European Integration: A Review Article”, in *Journal of Common Market Studies*, Vol. 37, No. 2, 1999, pp. 317–331; F. Laursen, “Theory and Practice of Regional Integration”, in *Jean Monnet/Robert Schuman Paper Series*, Vol. 8, No. 3, 2008; F. Schimmelfennig, “Regional Integration Theory”, in *Oxford Research Encyclopedia of Politics* (pre-publication version of article), Oxford University Press, 2018; F. Schimmelfennig, “European Integration (Theory) in Times of Crisis. A Comparison of the Euro and Schengen Crises”, in *Journal of European Public Policy*, Vol. 7, No. 25: 969–89, 2018.

<sup>②</sup> 事实上，莫拉维斯克早在1993年就提出了自由政府间一体化的理论雏形。参见 A. Moravcsik, “Preferences and Power in the European Community: A Liberal Intergovernmental Approach”, in *Journal of Common Market Studies*, Vol. 31, No. 4, 1993, pp. 473–523.

<sup>③</sup> L. Hooghe and G. Marks, “Is Liberal Intergovernmentalism Regressive? A Comment on Moravcsik (2018)”, in *Journal of European Public Policy*, 2019.

<sup>④</sup> T. Gehring, “Integrating Integration Theory: Neo-functionalism and International Regimes”, in *Global Society*, Vol. 10, No. 3, 1996, pp. 225–253.

尊重<sup>①</sup>。在此背景下，国家间一体化研究领域迄今没有出现被各方广泛接受的一般性理论。当然，莫拉维斯克并不承认这一点<sup>②</sup>。在他看来，自由政府间主义的三阶段博弈框架提供了新功能主义和后功能主义一体化理论所缺乏的参与者行为微观解释，可为进一步研究国家间一体化提供不可或缺的理论基础，本身就属于可整合新功能主义和后功能主义一体化理论的“大理论”（Grand Theory）。莫拉维斯克的上述观点，不出意料地招致了胡格等后功能主义一体化理论倡导者的激烈批评。胡格等人认为自由政府间主义和新功能主义、后功能主义一体化理论一样，只是国家间一体化现象的一个理论视角，而非可全面解释国家间一体化现象的“大理论”<sup>③</sup>。在胡格等人看来<sup>④</sup>，为解释2008年以来欧洲一体化困境，莫拉维斯克一方面强调欧债危机、移民危机、英国脱欧、东西欧分歧凸显等问题的严重性，另一方面从国家偏好形成中的经济因素拓展到强调社会、政治、宗教等其他因素，既不符合当前欧洲一体化面临的现实，又削弱了自由政府间主义自身的理论规范性和解释能力。

尽管彼此争论不休，但无论自由政府间主义倡导者还是新功能主义、后功能主义一体化理论倡导者，都承认其竞争对手的主张有其合理性，至少可对国家间一体化现象提供部分解释。譬如，莫拉维斯克写道，新功能主义者、后功能主义者“值得赞誉，不仅因为他们发现了令人迷糊的新经验趋势，也因为他们重新激起了关于一体化根本动力的理论争论”<sup>⑤</sup>。胡格等人也承认，自由政府间主义和新功能主义、后功能主义一体化理论一样“不存在需要被彻底抛弃的根本性错误”<sup>⑥</sup>。多数学者持有与此相似的观点，自由政府间主义、新功能主义和后功能主义一体化理论并非是相互对立的，而是相互补充的。对于格林而言，自由政府间主义、新功能主义和后功能主义一体化理论是

---

① D. J. Puchala, “Institutionalism, Intergovernmentalism and European Integration: A Review Article”, in *Journal of Common Market Studies*, Vol. 37, No. 2, 1999, pp. 317 – 331.

② A. Moravcsik, “Preferences, Power and Institutions in 21st-century Europe”, in *Journal of Common Market Studies*, Vol. 56, No. 7, 2019, pp. 1468 – 1674.

③ L. Hooghe and G. Marks, “Grand Theories of European Integration in the Twenty-first Century”, in *Journal of European Public Policy*, Vol. 26, No. 8, 2019, pp. 1113 – 1133.

④ L. Hooghe and G. Marks, “Is Liberal Intergovernmentalism Regressive? A Comment on Moravcsik (2018)”, in *Journal of European Public Policy*, 2019.

⑤ A. Moravcsik, “Preferences, Power and Institutions in 21st-century Europe”, in *Journal of Common Market Studies*, Vol. 56, No. 7, 2019, pp. 1468 – 1674.

⑥ L. Hooghe and G. Marks, “Grand Theories of European Integration in the Twenty – first Century”, in *Journal of European Public Policy*, Vol. 26, No. 8, 2019, pp. 1113 – 1133.

“一枚硬币的两面”“对同一现象不同角度的观察”<sup>①</sup>。唐纳德·帕查拉则认为，国家间一体化理论的未来挑战在于“将制度主义和政府间主义整合成某种形式的统一理论”<sup>②</sup>。然而，自由政府间主义和新功能主义、后功能主义一体化理论不只是“一枚硬币的两面”，同时也是对于不同类型的国家间一体化的不同解释。因此，构建可全面解释国家间一体化现象的一般理论，不仅要能够提供自由政府间主义和新功能主义、后功能主义一体化理论所能提供的视角，还要能为不同类型的国家间一体化现象提供一致有效的解释。

本文旨在对自由政府间主义、新功能主义、后功能主义一体化理论进行统合，构建可全面解释国家间一体化现象的一般理论。为此，首先需要从观察的视角和关注的类型两方面对自由政府间主义和新功能主义、后功能主义一体化理论进行比较，藉以对本文的研究视角和研究对象做进一步刻画。(1) 观察视角。自由政府间主义的视角是静态的，主要关注国家间一体化的动力、一体化进程中的关键参与者、国家愿意让渡和汇聚主权的条件<sup>③</sup>。新功能主义、后功能主义一体化理论的视角是动态的，主要关注国家间一体化演进/退化的过程，而不太考虑一体化的动力以及主权让渡和汇聚的条件，并假定非国家参与者在一体化中扮演着重要的角色。换句话说讲，自由政府间主义强调的是关于特定议题相关各方的博弈行为和博弈结果，而新功能主义、后功能主义一体化理论更强调特定议题博弈结果的“溢出/回溢效应”，以及路径依赖/自我挫败<sup>④</sup>。(2) 一体化类型。按照所遵循的不同国家关系学说，国家间一体化可分为理性主义一体化和理想主义一体化。两类一体化都拒绝接受现实主义国际关系理论，都相信国家之间存在结构化合作互利的可能，视国家之间结构化合作为解决跨国性的重要手段<sup>⑤</sup>。不同的是，前者被看作是成员国家实现自身利益的工具，后者则旨在通过消除民族主权国家的手段追求所有成员的共同利益。从假设来看，自由政府间主义、后功能主义一体化理

<sup>①</sup> T. Gehring, “Integrating Integration Theory: Neo-functionalism and International Regimes”, in *Global Society*, Vol. 10, No. 3, 1996, pp. 225 - 253.

<sup>②</sup> 帕查拉在文中将功能主义、新功能主义归类为制度主义，即强调制度对于社会的建构作用。参见D. J. Puchala, “Institutionalism, Intergovernmentalism and European Integration: A Review Article”, in *Journal of Common Market Studies*, Vol. 37, No. 2, 1999, pp. 317 - 331.

<sup>③</sup> A. Moravcsik, “Preferences, Power and Institutions in 21st-century Europe”, in *Journal of Common Market Studies*, Vol. 56, No. 7, 2019, pp. 1468 - 1674.

<sup>④⑤</sup> L. Hooghe and G. Marks, “Grand Theories of European Integration in the Twenty-first Century”, in *Journal of European Public Policy*, Vol. 26, No. 8, 2019, pp. 1113 - 1133.

论观察的是理性主义国家间一体化，新功能主义观察的是理想主义国家间一体化。在莫拉维斯克眼里，所有一体化都是理性主义一体化，都只是成员国追求自己利益的手段。对于胡格等而言，一体化演进还是退化取决于参与者国内自决意识是否足够强烈。新功能主义植根于多元主义和功能主义，将国家一体化视为以建立联邦为目的的过程<sup>①</sup>。然而在现实世界中，两类国家间一体化是并存的<sup>②</sup>。故此，可涵盖自由政府间主义、新功能主义和后功能主义一体化理论的国家间一体化大理论，既要解释理想主义一体化也要解释理性主义一体化，既要有效解释欧洲一体化现象也要逻辑一致地解释东南亚、非洲、南美洲、加勒比等所有地区的国家间一体化现象<sup>③</sup>。

为同时纳入动态和静态观察视角，既满足刻画理性主义国家间一体化的需要，又满足刻画理想主义国家间一体化的需要，本文将借鉴国家建构理论思路，对莫拉维斯克的三阶段博弈模型进行改造，构建多次重复的两阶段国家间一体化博弈模型，尝试借此对自由政府间主义、新功能主义和后功能主义一体化理论进行整合，以构建关于国家间一体化的一般理论。接下来本文将分为四个部分论述：第一部分讨论国家间一体化的概念及其测度，回答什么是国家间一体化以及如何测度国家间一体化；第二部分将构建多次重复的两阶段博弈模型，着重讨论国家偏好的形成；第三部分分析博弈均衡结果，包括单次博弈结果和重复博弈结果；第四部分是对全文的总结，并指出了本文理论模型的不足之处和未来的研究方向。

---

① T. Gehring, “Integrating Integration Theory: Neo-functionalism and International Regimes”, in *Global Society*, Vol. 10, No. 3, 1996, pp. 225–253.

② 欧共体之父多属于理想主义者，建立欧洲联邦是他们推动欧洲一体化的最终目标。但是，英国参与欧洲一体化更多是出于对自身利益的考量，而非寻求欧洲的整体利益。而且，国家间一体化不仅仅是欧洲现象，在其他大洲也都存在一体化现象。而这些地区的国家一体化的目的显然与欧洲国家不同，通过一体化寻求自身安全和经济利益是其基本出发点。

③ 格林于1996年曾问道，“如果认为欧洲共同体是一个经验上独特的现象，它真的需要自己的理论吗？如果现代新功能主义确实是传统政府间主义以外的现象，那么欧洲共同体是唯一可以以这种方式（新功能主义）加以研究的经验问题吗？”参见 T. Gehring, “Integrating Integration Theory: Neo-functionalism and International Regimes”, in *Global Society*, Vol. 10, No. 3, 1996, pp. 225–253；多森罗德在2010年尝试整合联邦主义理论和新功能主义时，其目的也是要建立一个可解释欧洲、非洲、亚洲等所有地区一体化现象的一般理论，而非仅仅关注欧洲一体化问题。参见 S. Dosenrode, “Federalism Theory and Neo-Functionalism: Elements for an Analytical Framework”, in *Perspectives on Federalism*, Vol. 2, Issue 3, 2010.

## 一 国家间一体化及其测度

本节我们将回答什么是国家间一体化，又如何测度国家间一体化。这是构建国家间一体化理论绕不过去且必须首先回答的两个问题。提出这两个问题，对从事国家间一体化研究的学者而言多少有些尴尬，历经半个多世纪竟然对此最基本的问题还没有一个公认的答案<sup>①</sup>。

### （一）什么是国家间一体化

通过经典文献梳理，我们会发现关于国家间一体化的定义大致可分为三类<sup>②</sup>。第一类，将国家间一体化视为一种过程。哈斯是其代表，他认为“通过这一过程，处于几个不同国家环境的政治行为体被说服将其忠诚、期望和政治行动转向一个新的中心，该中心的机构拥有或者要求拥有对既有民族国家的管辖权”；而这一过程的“最终目的是取代既有民族国家的新政治共同体的诞生”。<sup>③</sup> 第二类，视国家间一体化为一种状态。多伊奇是其代表，他认为一体化“意味着在成员之间达成一种‘共同体意识’并形成广泛且强有力的制度和惯例”，是“一种关系，在此关系之下政治行为者、单位或组成部分的行为是受到制约的，即不同于他们没有结合在一起的情况。……成员国可共同创造在彼此独立的情况下无法实现的系统财富”。<sup>④</sup> 对多伊奇而言，一体化并不必然导致一个类似于国家的新实体诞生。第三类，主张国家间一体化既是一种过程也是一

<sup>①</sup> 据马克卢普 1977 年考证，直到二战结束后，“一体化”一词才伴随着马歇尔欧洲重建计划的实施正式出现在官方文件和学术论文中。参见 F. Machlup, *A History of Thought of Economic Integration*, London: Macmillan, 1977, p. 3.

<sup>②</sup> S. Dosenrode, “Federalism Theory and Neo-Functionalism: Elements for an Analytical Framework”, in *Perspectives on Federalism*, Vol. 2, Issue 3, 2010.

<sup>③</sup> 哈斯关于国家间政治一体化目的的论断引起巨大争论，而且看起来很不现实。故此，另一位新功能主义倡导者林德伯格在 1963 年后退一步，将政治一体化定义为没有终点的过程：（1）各国放弃彼此独立执行外交政策和主要国内政策的愿望和能力，寻求做出联合决策或将决策过程委托给新的中央机构的过程；（2）说服不同环境下的政治参与者将他们的期望和政治活动转移到新的中心的过程。参见 Ernst B. Haas, *The Uniting of Europe: Political, Social, and Economic Forces 1950-1957*, Stanford: Stanford University Press, 1958; L. N. Lindberg, *The Political Dynamics of European Economic Integration*, Stanford: Stanford University Press, 1963, p. 6; 另外，马克卢普在 1977 年也将一体化视为目的的过程，只是观察的是经济一体化。他提出，经济一体化是将各个分散的小经济体合并成一个更大经济体的过程，其目的是“充分利用有效分工的所有潜在机会”。参见 F. Machlup, *A History of Thought of Economic Integration*, London: Macmillan, 1977.

<sup>④</sup> K. Deutsch, et al., *Political Community and the North Atlantic Area: International Organization in the Light of Historical Experience*, Princeton: Princeton University Press, 1957, pp. 1-91.

种状态。经济学家巴拉萨是其代表，她认为一体化“作为过程，它包含各种消除对属于不同主权国家经济体的歧视的举措；作为状态，它表现为各种形式的主权国家之间的歧视性待遇灭失”<sup>①</sup>。此外，政府间主义倡导者更愿意讨论国家间合作而非国家间一体化，霍夫曼、莫拉维斯克等经典作者都没有对国家间一体化作出明确定义。

尽管上述三类定义在某种程度上都属于盲人摸象，但还是抓住了国家间一体化的主要特征。沿着哈斯、巴拉萨等人的思路，本文将国家间一体化定义为，公共产品消费由覆盖单一主权国家扩大到由所有缔约主权国家结构化共享，公共产品供给由主权国家分别负责转为由主权国家结构化集体负责。公共产品消费覆盖由主权国家扩大到所有缔约各方，意味着制度的国家间边界被打破，所有成员方的所有行为主体都可以自由地跨边界开展该公共产品所支持的活动。公共产品提供责任由主权国家政府转移到国家间机构，意味着缔约各方的所有行为主体由向所在国家忠诚转为对所在的通过一体化形成的国家集团忠诚。按照制度学派假设，以文化和身份认同为基础的共同体为界，共同体成员对于共同体内部事务拥有集体自决的权力。在国家层面上，这一自决权力表现为国家主权。从这个意义上，本文定义的国家间一体化大致等同于成员国的主权让渡和汇聚。

与传统定义相比，本文上述定义可更全面反映国家间一体化的特征。首先，本文的定义同时涵盖了政治、经济两方面的一体化。公共产品既涉及经济又涉及政治，故此关于经济一体化和政治一体化的讨论可以在同一框架内进行。哈斯等政治学者关注政治行为体的忠诚转移，体现了政治的一面，但未能反映其经济的一面；巴拉萨等经济学者关注国家间边界的形成对跨国自由经济活动障碍的消除，体现了经济的一面，但未能对其政治的一面予以充分考虑。其次，本文主张国家间一体化兼具动态和静态双重属性。一体化的动态性意味着一体化可看作是一个过程，即公共产品缔约各方结构化共享、缔约各方结构化集体供应的促进/削弱；一体化的静态性意味着一体化可看作是一种状态，在该状态下公共产品在缔约各方之间结构化共享、由缔约各方结构化集体供应。显然，本文这一主张沿袭的是经济学家巴拉萨的观点。哈

<sup>①</sup> 相比较而言，卡纳特等人在1969年给出的定义与巴拉萨的定义较为接近，但表达相对简洁，即“经济一体化就是逐步消除在国家边界上发生的歧视的过程”。参见 Bela Balassa, *The Theory of Economic Integration*, London: Allen and Unwin, 1961, p. 1; F. Kahnert, et al., “Economic Integration Among Developing Countries”, Paris: Development Center, OECD, 1969.

斯强调一体化的过程属性，忽略了它的状态属性。多伊奇则相反，强调一体化的状态属性，忽略了它的过程属性。再次，本文认为国家间一体化是双向而非单向的<sup>①</sup>。哈斯将国家间一体化定义为有目的的过程，暗含着国家间一体化单向、不可逆的假定。巴拉萨在定义中虽未明确指出一体化的目的性，但她划分经济一体化五阶段时明显倾向于单向性假设。但现实中，既存在国家间一体化不断深化、拓展和扩大（即演进）的现象，还存在国家间一体化逐步浅化、收窄、缩减（即退化）的现象。在冷战结束后的欧洲，这一点表现得尤为明显。在其一体化进程中，既出现了许多退化的事件，如苏联和捷克斯洛伐克解体，英国和意大利进一步向地方分权，苏格兰和巴斯克地区的分离运动趋于活跃等；也出现了许多演进的事件，如欧盟由12国扩大为28国，欧元作为单一货币出现，德、法等国将货币主权转移给欧洲央行等<sup>②</sup>。最后，本文所定义的国家间一体化，不仅限于区域国家一体化，无须参与国家的国土彼此连接；对于参与国家的数量也无须三个以上的要求，两个国家之间同样存在一体化的问题<sup>③</sup>。

## （二）如何测度国家间一体化

2010年多森罗德沿着巴拉萨的思路将经济一体化划分成随机合作、自由贸易协定、关税同盟、共同市场、经济同盟、完全一体化6种形态，将政治一体化划分为5个政治一体化阶段。他这样划分有其道理和意义，但是远不够精细，不能对不同地区同一形态的一体化现象进行有效比较。譬如，按照关税及贸易总协定（GATT）给出的定义，截至目前全球有13个关税同盟，但显然并非所有关税同盟的发育程度都相同，如非洲联盟的发育程度显然比不上海湾合作委员会。本文将沿用多森罗德不区分国家间一体化与国家间合作的办法，但将国家间一体化看作是具有连续性变化、而非离散形态的光谱。其中，临时性、随机性的合作是光谱的一端，联合形成新的联邦国家是光谱的另一端。基于前面给出的定义，本文将沿着三个维度对国家间一体化的发

<sup>①</sup> 在这一点上，本文接受霍夫曼1966年的主张，认为国家之间存在促进一体化的引力，也存在推动彼此分离的张力。

<sup>②</sup> M. Ruta, "Economic Theories of Political (Dis) integration", in *Journal of Economic Surveys*, Vol. 19, No. 1, 2005.

<sup>③</sup> 施莫芬尼于2018年提出区域一体化组织应满足四个条件：（1）为国家所建立，以国家为成员；（2）有固定的办公地点和工作人员，有确定的办事和决策程序，有自主决策和实施能力；（3）有两个以上的成员；（4）成员在地理上彼此连接。参见F. Schimmelfennig, "Regional Integration Theory", in *Oxford Research Encyclopedia of Politics* (pre-publication version of article), Oxford University Press, 2018.

育状态进行测度。(1) 共享的公共产品范围。其一端是没有任何共享的公共产品，另一端是覆盖国家能提供的所有公共产品。共享的公共产品范围扩大，意味着国家间一体化功能上的拓展，反之则代表国家间一体化功能收窄。(2) 集体供给的公共产品比例。其一端是0，即所有的公共产品都由成员国单独提供；另一端是1，即所有的公共产品都由成员国集体提供。集体供给的公共产品比例上升，意味着国家间一体化程度的深化，反之则意味着国家间一体化程度浅化。(3) 共享公共产品的成员数量。共享公共产品的成员数量增加，参与公共产品集体供给的成员增多，意味着国家间一体化规模扩大，反之则意味着国家间一体化规模缩小。

我们用  $T = T(C, R, N)$  来描述国家间一体化、国家主权让渡和汇聚状态， $T$  值越大则国家间一体化、主权让渡和汇聚水平越高。其中， $C$  为共享的公共产品范围变量，取值范围为  $[0, 1]$ 。 $C = 0$  表示共享的公共产品集合为空， $C = 1$  表示共享的公共产品集合等同于国家提供的公共产品集合。 $R$  为集体供给的公共产品比例，根据前面的界定确定取值范围为  $[0, 1]$ 。 $N$  为共享公共产品的成员数量，取值范围为  $[2, N_m]$ ，其中  $N_m$  表示现有独立主权国家的数量。对于任意  $N \geq 2$ ，存在两个极端状态。其一端， $C = 1, R = 1$ ，表示  $N$  个主权国家实现完全一体化，形成了新的国家；另一端， $C = 0, R = 0$ ，表示  $N$  个主权国家之间只存在临时性、随机性的合作。根据上一段的界定，令  $\frac{\partial T}{\partial C} > 0$ ， $\frac{\partial T}{\partial R} > 0$ ， $\frac{\partial T}{\partial N} > 0$ 。即，国家间一体化在功能上不断拓展，在程度上不断深化，在规模上不断扩大，则越向形成联邦国家的一端靠近，国家主权让渡和汇聚越强，意味着国家间一体化演进；反之，国家间一体化功能收窄，程度浅化，范围缩减，则向随机性、临时性合作的一端靠近，国家主权让渡和汇聚减少，意味着国家间一体化退化。

## 二 多次重复的两阶段博弈模型

本节我们将借鉴国家建构理论思路对莫拉维斯克三阶段博弈模型进行改造，构建多次重复的两阶段国家间一体化博弈模型。模型构建过程中，我们既关注国家间一体化的动力、关键参与者、国家主权汇聚与让渡的静态条件，又关注国家间一体化演进/退化的动态过程。

### (一) 基本假定

首先, 本文将考察两个国家之间的一体化。将国家间一体化博弈简化为双边博弈, 是因为双边博弈处理起来相对简单。而多方博弈即使三方模型都过于复杂, 难以得出有价值的分析结论。更重要的是, 将国家间一体化博弈简化为两方博弈并不失一般性。现实中, 有效果、有效率的国家之间的谈判都是双边谈判, 多边谈判实质上也被分解成一系列双边谈判。在多边谈判中, 通常最积极或者最有权力的两方首先进行谈判, 达成一致后再作为一个整体与其他参与方进行谈判。这样, 多边谈判中任何参与方面对的都是单一的谈判对手。譬如, 刚刚结束的美墨加贸易谈判, 美国和墨西哥首先达成协议, 然后作为一个整体与加拿大进行谈判<sup>①</sup>。事实上, 联合国、世界贸易组织等全球性多边体系下的谈判也都采取类似的方式。譬如, 世贸组织多哈回合农业协定谈判中, 主要的发展中大国彼此协调立场结成“二十国集团”以一个声音说话, 与美国、欧盟和日本等发达经济体组成的联盟进行谈判, 试图争取自己的最大利益<sup>②</sup>。由此, 给定国家间一体化参与方包括 A、B 两个国家, 国家间一体化状态值函数被进一步简化为  $T = T(C, R)$ 。

其次, 本文假定共享的公共产品范围变量值与公共产品的集体供给比例一一对应。从方式来看, 一体化可以是整合也可以是融合。对于整合型国家间一体化, 合成一体的各方中至少有一方是一体化的被动接受者。由于要求对所涉对象进行调整, 整合型国家间一体化往往意味着某种形式的暴力强制。对于融合型国家间一体化, 合成一体的各方中没有被动接受者。融合型以自由缔结契约为基础, 即使存在调整改变的需要也通常以自愿为原则。历史上, 国家间的一体化多属于整合型一体化<sup>③</sup>。本文关注的是二战以来发端于西欧并遍及北美、亚、非、拉地区的独立主权国家自发自愿的融合型一体化实践。对这类融合型的国家间一体化, 公共产品单独供给和公共产品完全共享, 或者公共产品全部集体供给而公共产品共享集合为空的情形基本不会出现。在

<sup>①</sup> 再譬如, 英国脱欧谈判中, 德国、法国及其他欧盟国家虽然都是利益相关方, 但并不直接与英国进行谈判, 而是以欧盟的名义作为一个整体与英国谈判。

<sup>②</sup> “二十国集团”是由 20 个主要发展中国家组成的代表发展中国家利益的非正式谈判联盟, 在 2003 年 8 月世贸组织第五次部长级会议(坎昆会议)的会议筹备期间逐步形成。其标志是, 2003 年 20 国代表签署了一份针对欧盟和美国此前提出的农产品提案的替代性框架文件。二十国集团的成员国包括中国、印度、巴西、南非、阿根廷、墨西哥、巴基斯坦、泰国、委内瑞拉、菲律宾、巴拉圭等一些发展中国家。

<sup>③</sup> 譬如, 罗马帝国、蒙古帝国等古代帝国都是建立在军事暴力基础之上。

二战后席卷全球的国家间一体化实践中，公共产品集体提供比例通常与共享的公共产品的范围高度相关。譬如，1999年1月1日德国、法国等11个国家共享欧元作为主权货币的同时，也将货币政策制定的权力让渡给了欧洲央行。为分析简便，本文假定 $C$ 与 $R$ 一一对应，即给定共享的公共产品范围对应着唯一确定的集体供给比例，同时给定公共产品集体供给比例对应唯一确定的共享的公共产品范围。由此，国家间一体化状态值函数被进一步简化为 $T = T(C)$ 。

再次，本文对应三种国际关系理论将民众对于国家间一体化的偏好分为三类。在这里，民众的偏好是指民众对于国家所追求的价值价值的认可，包括保障和维护国家主权统一、维系与其他国家之间的和平、促进经济增长、实现文化繁荣、保护生态环境等。其中，保障和维护国家主权统一可视为主权让渡减少，用 $T$ 表示；促进经济增长目标，用 $E$ 表示；维系与其他国家之间的和平、实现文化繁荣、保护生态环境等价值目标，用 $O$ 表示。相应地，民众眼中的国家利益是促进和保障国家所追求的价值价值的实现。对现实主义者而言，生存是国家第一且唯一的要务，而所谓生存就是保障和维护国家主权的统一，为此不惜牺牲经济利益、国家间的和平以及所有其他目标<sup>①</sup>。与此对应，现实主义者的效用函数 $U_R = U(T)$ ，且 $\frac{\partial U_R}{\partial T} < 0$ 。对理性主义者而言，保障和维护国家主权统一是国家的重要偏好但非唯一偏好，维系国家间和平、促进经济增长、实现文化繁荣、保护生态环境同等重要。因此，理性主义者的效用函数 $U_I = U(T, E, O)$ ，且 $\frac{\partial U_I}{\partial T} < 0$ ， $\frac{\partial U_I}{\partial E} > 0$ ， $\frac{\partial U_I}{\partial O} > 0$ 。对于理想主义者而言，维护国家间和平、保护生态环境等是最重要的国家利益，为此甚至可以牺牲国家主权和经济利益。由此，理想主义者的效用函数 $U_F = U(O)$ ，且 $\frac{\partial U_F}{\partial O} > 0$ 。

## （二）模型框架

本文模型构建借鉴了国家形成模型的思路。国家形成模型最初由阿尔塞纳和斯波洛雷共同提出，其基本想法是国家类似于俱乐部，人们结构成为一

<sup>①</sup> 需要提及的，在国际关系问题上，民众的偏好往往因涉及的对象不同而存在差异。譬如，法国国民通常会以理想主义方式看待与德、意、荷、比、卢五国的关系，以理性主义方式对待英国等其他欧盟国家及美国，但对俄罗斯更大可能会采取现实主义态度。另外，一国与另一国的交往方式还随着国际环境的变化而变化。譬如，英德在二战期间是敌对关系，但战后两国交往更多基于理性主义立场。

个国家（共同体），因为国家能够以低成本提供诸如防卫、法律和秩序、货币等公共产品<sup>①</sup>。其后，鲁塔沿着这一思路对国家间一体化以及去一体化的动机和激励因素进行了分析<sup>②</sup>。类似地，本文把国家间一体化视为俱乐部形成，主张各国愿意结构成为一个国家俱乐部，因为俱乐部可以降低公共产品供给成本，增加其他方面如经济、安全等方面的额外收益。不同的是，（1）国家形成博弈中的参与主体是公民个体或利益集团，国家俱乐部形成博弈中的参与主体则是国家。（2）阿尔塞纳等人关注的是国家间一体化的最优成员规模问题，关注参与国家在一体化成员数量扩大带来的规模经济收益与异质化导致的弥合成本增加之间的权衡。本文除分析一体化成员数量规模扩大对于博弈过程 and 结果的影响，还重点关注国家在维护国家主权独立与主权让渡收益之间的权衡。（3）阿尔塞纳等人将国家关于一体化的偏好视为给定的条件，国家在一体化问题上关注的重点是经济利益；本文模型则认为国家偏好的形成是各国对国家价值认同不一的个体或集团相互博弈的结果，其中既有重视经济利益者如制造商集团，也有强调国家自决、反对让渡主权的个体或集团。

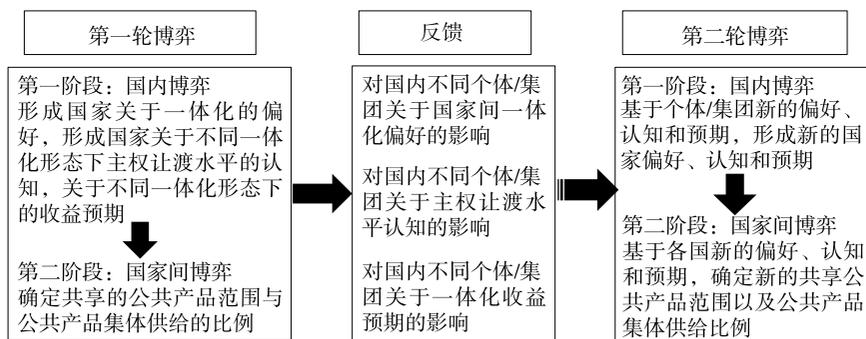


图1 国家间一体化博弈模型结构图

资料来源：笔者绘制。

本文将国家间一体化视为多次重复的博弈过程，每一轮博弈都包含国内

<sup>①</sup> 阿尔塞纳等人于2000年讨论了经济一体化与政治逆一体化的关系，提出贸易制度变化会影响国家形成，其落脚点也是最优国家规模。参见 A. Alesina and E. Spolaore, "On the Number and Size of Nations", in *Quarterly Journal of Economics*, Vol. 112, No. 4, 1997, pp. 1027-1056; A. Alesina, E. Spolaore, and R. Wacziarg, "Economic Integration and Political Disintegration", in *American Economic Review*, Vol. 90, No. 5, 2000, pp. 1276-1296.

<sup>②</sup> M. Ruta, "Economic Theories of Political (Dis) integration", in *Journal of Economic Surveys*, Vol. 19, No. 1, 2005.

博弈和国家间博弈两个阶段，前一轮博弈结果通过反馈机制对下一轮博弈产生影响（见图1）。在国内博弈阶段，参与方为两国的公民个体/利益集团，博弈结果是两国关于国家间一体化的国家偏好形成，关于国家间一体化带来的主权让渡水平认知形成，关于国家间一体化的收益预期形成。在国家间博弈阶段，参与方为代表两国国家利益和偏好的政府，博弈结果是确定两国共享的公共产品范围与公共产品集体供给比例。随着共享的公共产品范围与公共产品集体供给比例确定，两国一体化/逆一体化后果逐步显现，从而影响国内不同个体/集团关于国家间一体化偏好、关于主权让渡水平的认知，以及一体化收益预期。一旦这种影响突破某一临界点，就会产生新的国家间一体化/逆一体化需求，诱发新一轮国家一体化博弈。新一轮博弈同样包含国内博弈和国家间博弈两个阶段。在国内博弈阶段，基于两国公民个体/利益集团新的偏好、认知和预期，形成两国新的国家偏好、认知和预期。在国家间博弈阶段，基于两国新的国家偏好、认知和预期，确定两国新的共享公共产品范围以及公共产品集体供给比例。“博弈→反馈→博弈”的过程不断重复，两国一体化由此不断深化或者浅化、拓展或者收窄，直到达到某种稳态均衡。

与莫拉维斯克一样，本文模型中国家间博弈的参与主体是作为国家代表的政府。但这并不意味着本文否认超国家组织、跨国家组织、次国家组织在国家间一体化博弈中的重要性，相反，本文承认在很多情形下它们甚至是国家间一体化最重要的推动者。本文做此处理，是因为现实中各国政府是（至少名义上是）国家间谈判的主体、国家间条约的签订者，拥有（至少名义上拥有）对一体化组织或机构的所有权利。在此背景下，超国家组织、跨国家组织、次国家组织在国家间一体化博弈中的推动作用，能且只能通过影响参与国家的偏好、进而影响博弈进程和结果来实现，在本文模型中则反映在每轮博弈的国内偏好形成阶段以及在反馈环节对公民个体/利益集团偏好、认知和预期的影响上。除此之外，本文模型与莫拉维斯克理论的不同之处有三。其一，本文模型中的每轮博弈只有国内博弈和国家间博弈两个阶段。在莫拉维斯克的框架中，国家间主权让渡博弈被分成两阶段，第一阶段确定主权让渡内容以及参与方各自的收益，第二阶段确定保障让渡和各方利益的机制。但按照本文前面给出的“共享的公共产品范围变量值与公共产品的集体供给比例一一对应”的假定，在确定一体化预期收益分配的同时也确定了一体化所需的主权让渡安排。其二，本文模型中国民/国家的偏好是总偏好的概念。在莫拉维斯克看来，国民/国家关于国家间一体化偏好是基于具体事项的，国

民之间和国家之间需要就每一特定议题进行博弈<sup>①</sup>。本文则认为国民考虑的是个人总效用，国家考虑的所有国民的总福利，故允许国民/国家就影响效用/福利的所有议题同时展开博弈。其三，本文模型是重复博弈而非一次博弈模型，两次博弈之间存在反馈环节。本文引入反馈环节除试图反映超国家组织、跨国家组织、次国家组织对国家间一体化博弈的推动作用外，还试图借此描述新功能主义倡导者所强调的因路径依赖带来的“溢出效应”，以及后功能主义者所强调的国际国内政治环境变化带来的“回溢”现象。本文采用多次重复博弈的模型架构，旨在借此反映国家间一体化的过程性和动态性，揭示国家间一体化演进或者退化过程背后的动力与机制。

### 三 博弈均衡分析

本节将讨论国家间主权汇聚与让渡的动力和条件，以及一体化演进/退化的反馈机理。主要内容包括：（1）沿着霍特林空间竞争模型思路分析参与国国家偏好行为；（2）在古诺博弈框架下分析参与国主权让渡与红利分享比例的均衡关系；（3）探讨国家间一体化反馈机制；（4）分析9种不同反馈组合下参与国的稳态均衡选择；（5）讨论参与国数量变化对均衡选择的影响。

#### （一）第一轮博弈

假定初始状态下，国家A和B的交往完全基于现实主义理念，均将生存即保障和维护主权统一视为唯一要务，将对方视为行为不可预知者。基于这样的理念，A和B之间只有临时性的合作，没有任何结构化合作安排。即A和B均拥有完整的主权，在任何功能领域都不存在一体化，更谈不上集体决策的比重了<sup>②</sup>。由于受到某种外力作用造成扰动，A、B两国国民偏好发生变化，过去一致同意采取现实主义逻辑处理双边关系，如今两国都出现了理性

<sup>①</sup> 莫拉维斯基的原模型博弈支付只考虑经济收益，“一体化偏好主要来自于强有力的制造商集团的商业利益”，“政府将一体化作为在预算和管制约束条件下追求制造商集团商业利益的手段”。但莫拉维斯基在其最新著作中放松了这一假定，不再否认身份认同等非物质偏好对于国民/国家偏好的影响。不过，莫拉维斯基并没有就如何把这些非物质偏好纳入博弈模型展开讨论。参见 A. Moravcsik, *The Choice for Europe*, Ithaca: Cornell University Press, 1998; A. Moravcsik, “Preferences, Power and Institutions in 21st - Century Europe”, in *Journal of Common Market Studies*, Vol. 56, No. 7, 2019, pp. 1468 - 1674.

<sup>②</sup> 需要提及的是，这一初始状态假定并非必须是必须的，只是为了叙述方便。事实上，国家间博弈的初始状态可以是临时性、随机性的合作与联合形成新的联邦国家之间的任何一种状态，不影响模型分析框架的有效性。如果假定为A和B属于同一国家S的两个子共同体，将要讨论的问题是逆一体化过程，即S解体为两个国家A、B的过程。

主义和理想主义呼声。而且这样的呼声足够强大，迫使 A、B 两国政府不得不寻求一体化合作。博弈由此拉开序幕。

博弈第一阶段，A、B 两国国民个体/利益集团通过博弈形成国家偏好。假定国家偏好的形成遵循多元主义范式，每个群体的意志都可以得到自由表达。尽管“更强悍、更确定和组织的制度化程度更高的”利益诉求可能表达更为充分，而“越是孱弱、不清晰的”政策表达越不充分<sup>①</sup>，但国家偏好最终显示为赢得多数民众支持的政党的偏好。给定上述假定，我们将沿着霍特林空间市场竞争模型的思路，构建两党选举制下的国家偏好形成博弈模型<sup>②</sup>。

A 国国家偏好。进一步假定，(1) A 国对 B 国三类国际关系理念都有程度不同的支持者，在  $U_0 = U(T)$  和  $U_1 = U(O)$  两种极端情形中间连续分布，其中间偏好  $U_m = U(T) + U(E_A^e) + U(O_A^e)$ 。给定密度函数  $f(U)$ ，有  $\int_{U_0}^{U_1} f(U) dU = 1$ 。(2) A 国存在两个政党 a 和 b，其中政党 a 的政策  $U_a$  偏近于  $U_0$ ，且  $U_a \in [U_0, U_m]$ ；政党 b 的政策  $U_b$  偏近于  $U_1$ ，且  $U_b \in [U_m, U_1]$ 。(3) 选民倾向于投票给政策建议与自己偏好最相近的政党。给定偏好为  $U$  的选民，如果  $|U - U_a| \leq |U - U_b|$ ，那么该选民投票给政党 a 的概率  $P^a(U, U_a) = q(|U - U_a|)$ ，且  $\frac{dP^a}{d(|U - U_a|)} < 0$ ；投票给政党 b 的概率  $P^b(U, U_b) = 0$ 。如果  $|U - U_a| > |U - U_b|$ ，该选民投票给政党 a 的概率  $P^a(U, U_a) = 0$ ，投给政党 b 的概率  $P^b(U, U_b) = q(|U - U_b|)$ ，且  $\frac{dP^b}{d(|U - U_b|)} < 0$ 。由此，得到政党 a 和政党 b 的支持度函数分别为：

$$P(U_a) = \int_{U_0}^{U_a} q(U_a - U) * f(U) dU + \int_{\frac{U_a + U_b}{2}}^{U_a} q(U - U_a) * f(U) dU$$

$$P(U_b) = \int_{U_b}^{U_1} q(U - U_b) * f(U) dU + \int_{\frac{U_a + U_b}{2}}^{U_b} q(U_b - U) * f(U) dU$$

给定上述政策支持度函数，政党 a 和政党 b 将选择最优政策以实现支持度最大化。政党政策博弈的纳什均衡满足：

① A. Moravcsik, *The Choice for Europe*, Ithaca: Cornell University Press, 1998.

② 唐斯 (Anthony Downs) 是应用霍特林空间市场模型思想分析党派意识形态的先行者，其名著《民主的经济理论》第八章就此问题进行了详细讨论。

$$U_a^* = \text{Argmax}P(U_a^*, U_b)$$

$$U_b^* = \text{Argmax}P(U_a, U_b^*)$$

显然，上述纳什均衡解取决于A国国民偏好分布。不失一般性，为简化分析，假定A国国民偏好呈单峰分布<sup>①</sup>。

(1) 正态分布。如果A国国民偏好分布为正态分布，如图2-1曲线AA所示，那么纳什均衡解为  $U_a^* = U_b^* = U_m$ ，且  $P^{a*} = P^{b*}$ 。这意味着政党a和政党b都有机会当选，而且无论哪个政党当选最终表现出来的偏好都是中间选民的偏好<sup>②</sup>。由此可得  $U_A^* = U_m$ ，满足  $\frac{\partial U_A}{\partial T_A} \leq 0$ ， $\frac{\partial U_A}{\partial E_A^e} \geq 0$ ， $\frac{\partial U_A}{\partial O_A^e} \geq 0$ ，主权转让越少，经济与其他方面的预期收益越大则效用越高。为分析简便，本文假设其他非经济性收益可以按某种方式以经济收益加以度量<sup>③</sup>。令  $O_A^e = O(E_A^e)$ ，A国偏好相应简化为  $U_A = U_A(T_A, E_A^e)$ 。由此得到A国无差异曲线  $T_A = T_A(E_A^e)$ （见图3第一象限）， $\frac{\partial T_A}{\partial E_A^e} > 0$ ， $\frac{\partial^2 T_A}{\partial (E_A^e)^2} > 0$ 。即只有预期能得到更大的收益，A国才愿意让渡更大的主权，而且主权让渡所需的边际补偿递增（如图2-2所示，政策无差异曲线下凸）。

(2) 右偏分布。如果A国国民偏好为右偏分布（如图2-1曲线CC所示），偏差系数  $Cs_U > 0$ ，那么  $(U_a^*, U_b^*) = (U_{e1}, U_m)$ ，其中  $U_{e1} \in [U_0, U_m)$ ，且满足  $P^{a*} > P^{b*}$ 。也就是说，政党b仍旧选择中间偏好，而政党a为赢得选举将向右倾斜。 $P^{a*} > P^{b*}$  意味着政党a赢得了选举，故A国的国家偏好体现为政党a的偏好。即， $U_A^* = U_a^*$ ，满足  $\frac{\partial U_a^*}{\partial T} \leq 0$ ， $\frac{\partial U_a^*}{\partial E_A^e} \geq 0$ 。 $U_a^* < U_m$ ，意味着相较于正态分布情形，右偏分布情形下A国政策无差异曲线  $T_A^a = T_A^a(E_A^e)$  更靠近横轴，更远离纵轴。如图2-2所示，给定相同的红利预期，如果国民偏好呈右偏分布，则对应的主权让渡意愿弱于正态分布情形。

① 至于其他形态国民偏好分布下的国家偏好形成博弈，同样可采取本文给出的分析框架，只是均衡结果存在差异。其中，在国民偏好呈双峰分布情形下，政党政策或者左倾或者右倾，不会收敛于中间偏好。如果左峰更高，则左倾政党将赢得选举。如果右峰更高，则右倾政党赢得选举。如果双峰对称，则意味着国民在一体化问题上高度分裂，哪个政党都有机会执政，但都难以维系长期政策稳定。一旦出现政党交替，就会出现大幅度政策调整。关于多峰分布和均衡分布情形下的国家偏好形成博弈，唐斯的政党意识形态分析思路有一定的借鉴意义。

② 这一特殊情形与唐斯的结论一致，在两党体制中政党的意识形态收敛于中点。

③ 譬如，国家安全收益就可用一体化带来的国防支出减少加以度量。

(3) 左偏分布。如果 A 国国民偏好为左偏分布（如图 2-1 曲线 BB 所示），偏差系数  $Cs_U < 0$ ，那么  $(U_a^*, U_b^*) = (U_m, U_{e2})$ ，其中  $U_{e2} \in [U_m, U_1)$ ，且满足  $P^{a*} < P^{b*}$ 。换句话说讲，政党 a 仍旧选择中间偏好，而政党 b 为赢得选举而出现左倾。 $P^{a*} < P^{b*}$  意味着政党 b 赢得了选举，故 A 国的国家偏好体现为政党 b 的偏好。即， $U_A^* = U_b^*$ ，满足  $\frac{\partial U_A}{\partial T_A} \leq 0$ ， $\frac{\partial U_A}{\partial E_A^e} \geq 0$ 。 $U_b^* > U_m$ ，意味着相较于正态分布情形，左偏分布情形下 A 国政策无差异曲线  $T_A^b = T_A^b(E_A^e)$  更靠近纵轴，更远离横轴。如图 2-2 所示，给定相同的红利预期，如果国民偏好呈左偏分布，则对应的主权让渡意愿强于正态分布情形。

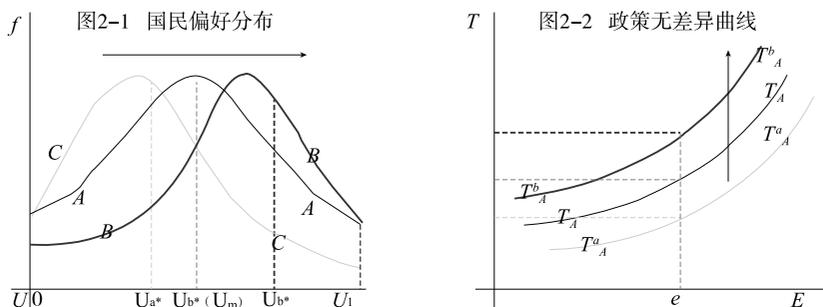


图 2 国民偏好分布与政策无差异曲线

资料来源：笔者绘制。

如图 2 所示，国民偏好从右偏分布到正态分布再到左偏分布（图 2-1 所示，从 CC 到 AA 再到 BB），国民偏好分布偏差系数  $Cs_U$  随之由正降为 0 再转为负，政策无差异曲线相应地向上由  $T_A^a$  移动至  $T_A$  再移动至  $T_A^b$ （图 2-2 所示）。更一般地，随着国民偏好分布偏差系数  $Cs_U$  下降，政策无差异曲线将向上移动，相同的红利预期对应着更大的主权让渡意愿。

B 国国家偏好。为分析的方便，本文关于 B 国的假设与 A 国假设完全对称，且同样只考虑国民偏好单峰分布的情形。由此，如果 B 国国民偏好呈正态分布，则其无差异政策偏好曲线为  $T_B = T_B(E_B^e)$ ， $\frac{\partial T_B}{\partial E_B^e} > 0$ 。与 A 国一样，B 国让渡更多主权的前提是得到更大的补偿，而且主权让渡所需的边际补偿递增。如果 B 国国民偏好呈右偏分布，则其政策偏好亦会左倾，无差异曲线较之正态分布情形靠下靠外。如果 B 国国民偏好呈左偏分布，则其政府偏好亦会右倾，无差异曲线较之正态分布情形靠上靠内。

博弈第二阶段, A 和 B 两国政府基于自身偏好和利益预期, 通过博弈确定主权让渡水平, 即确定共享的公共产品范围变量指标和公共产品集体供应比例。作为理性行为者, A 国、B 国政府愿意让渡主权以推进一体化的原因在于一体化带来的收益能弥补让渡主权造成的福利损失。为简化分析, 假定 (1) 国家间一体化的收益为公共知识, 且是 A 和 B 两国都接受的共识。根据前面的假定, 这里国家收益除经济收益外, 还包括以经济收益度量的诸如国家安全之类的其他利益。不失一般性, 令  $E_T^e = E_T(T)$ , 且  $\frac{\partial E_T^e}{\partial T} \geq 0$ ,  $\frac{\partial^2 E_T^e}{\partial T^2} < 0$ 。即, 假定 A 和 B 两国主权让渡越多则一体化预期收益越大, 且主权让渡的预期边际收益递减。相应地, 主权让渡红利曲线为上凸曲线, 如图 3 第一象限曲线  $\alpha_1 E(T)$ 、 $\alpha_2 E(T)$  所示。根据前面的定义, 国家间一体化意味着公共产品消费共享、公共产品集体提供。由此,  $T = T_A = T_B$ , 即一体化过程中 A 国和 B 国让渡的主权完全相同。(2) 国家间一体化的收益可以通过特殊的制度安排为 A 和 B 两国按比例分享<sup>①</sup>。由此, 令  $E_A^e = \alpha E_T^e$ ,  $E_B^e = (1 - \alpha)E_T^e$ , 分别为 A 国和 B 国一体化收益预期曲线。其中,  $\alpha$  为 A 国所得一体化红利分享比例。如果  $\alpha$  由  $\alpha_1$  上升至  $\alpha_2$ , 那么 A 国一体化收益预期曲线向外移动 (见图 3 第一象限), 同时 B 国一体化收益预期曲线向内移动 (见图 3 第二象限)。

联立方程  $T_A = T_A(E_A^e)$ ,  $E_T^e = E_T(T)$ ,  $T = T_A = T_B$ ,  $E_A^e = \alpha E_T^e$ , 得到 A 国的反应函数  $T_A = T_A(\alpha)$ , 且  $\frac{dT_A}{d\alpha} > 0$ 。即, 给定其他条件不变, A 国主权让渡意愿  $T_A$  与所得红利分享比例  $\alpha$  成正比 (图 3 第四象限红线所示)。联立方程  $T_B = T_B(E_B^e)$ ,  $E_T^e = E_T(T)$ ,  $T = T_A = T_B$ ,  $E_B^e = (1 - \alpha)E_T^e$ , 得到 B 国的反应函数  $T_B = T_B(\alpha)$ , 且  $\frac{dT_B}{d\alpha} < 0$ 。即, 给定其他条件不变, B 国主权让渡意愿  $T_B$  与红利分享比例  $\alpha$  成反比 (图 3 第三象限所示)。联立两个反应函数  $T_A = T_A(\alpha)$ 、 $T_B = T_B(\alpha)$ , 得到 A 和 B 两国一体化博弈的纳什均衡解  $(T^*, \alpha^*)$ 。即, 给定 A 和 B 两国关于国家间一体化的偏好, 给定国家间一体化预期收益函数形式, 主权让渡水平和红利分享比例在谈判中内生决定。图 3 中,  $(T^*, \alpha_1)$  为博弈均衡结果。即, 给定红利分享比例为  $\alpha_1$ , A 和 B 两国主权让渡

<sup>①</sup> 现实中, 国际贸易发展的一个主要障碍就是贸易红利难以在贸易伙伴间公平地分享。但也正因为红利分配难以做到各方满意的公平, 才会出现参与者事后对于事前约定的主权让渡及红利分配不满意, 由此诱发新一轮博弈。

意愿均为  $T^*$ 。如果红利分享比例由  $\alpha_1$  调整为  $\alpha_2$ ，那么 A 国主权让渡意愿由  $T^*$  上升至  $T_a$ ，B 国主权让渡意愿则由  $T^*$  下降至  $T_b$ ，原有均衡由此被打破。

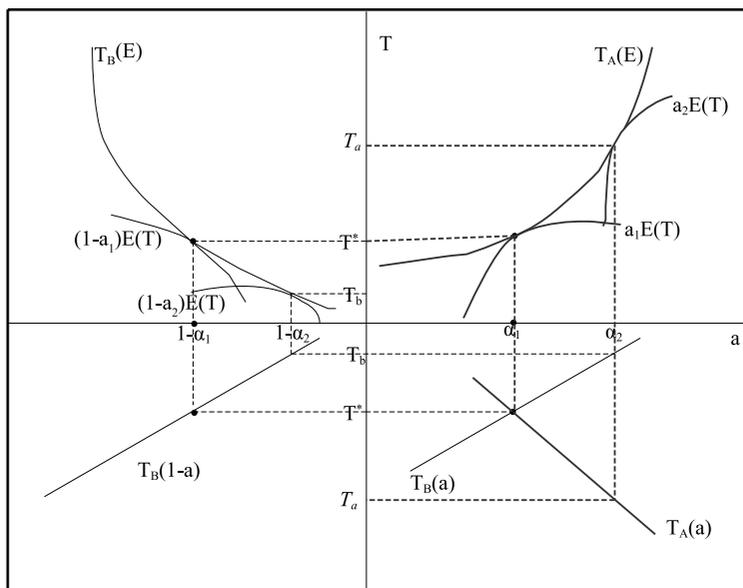


图3 国家间一体化博弈四象限图

资料来源：笔者绘制。

与此同时，均衡状态还将随两国偏好以及总收益预期而变化。给定总收益预期不变，相对于正态分布，两国偏好左倾，主权让渡意愿上升；两国偏好右倾，主权让渡意愿下降。给定其中一国偏好不变，另一国偏好左偏幅度越大，则均衡主权让渡意愿越低；另一国偏好右倾幅度越大，则均衡主权让渡意愿越高。给定两国偏好不变，一体化预期收益越大，则两国主权让渡意愿越强烈。一体化红利分享比例主要取决于两国偏好。如果两国偏好完全一致，即  $T_A(\cdot) = T_B(\cdot)$ ，那么红利分享比例  $\alpha$  确定为  $1/2$ ，两国平分预期红利。两国依存状态对于两国偏好有着很大影响。对于 A 国而言，单边可替代政策的价值越大，可替代联盟的价值越大，国家偏好左倾幅度就越大，同时政策无差异曲线会向右向下移动，就会要求分享更大比例的一体化红利；如果存在问题关联（issue-linkage）或者补偿性支付（side-payment）的机会，国家偏好就会向右倾斜，政策无差异曲线向左向上移动，就可能接受相对较小的红利分享比例。

## (二) 反馈

当两国谈判完成,就一体化红利分配比例、公共产品共享范围、公共产品集体供给比例达成一致后,两国就进入了相应的一体化状态。此后,反馈机制开始发挥作用,参与方国内不同个体/集团调整关于国家间一体化的偏好、关于主权让渡的认知、关于一体化红利的预期,逐步形成新的偏好、新的认知和新的预期。在此之前,我们先假定能找到一个合理的国内再分配机制,使得参与方从一体化中所得红利/所受损失可以公平地分配/转嫁给所有国民。换句话讲,在一体化问题上,个人得失与国家得失完全一致。故此,在国内偏好博弈形成过程中,个人只须考虑国家的得失而无须考虑自身的得失。另外,本文关于A和B国的假定是对称的,因此仅考虑一体化对A国国民的红利预期以及自决意识的影响。

反馈分成两大类6种情形。第一类反馈,A国国民关于一体化红利的预期变化,包括(1)(2)(3)3种情形。情形(1),国民主权让渡的红利预期上升。在图4中表现为,主权让渡的预期红利曲线 $E(T)$ 由 $E_1$ 向外(远离纵轴)移至 $E_2$ 。有两种情形会导致这一变化出现。其一,上轮一体化事后实现的红利大于事先预期的红利。在此情形下,参与方基于适应性预期规则,将上调其给定主权让渡不变情形下的未来红利预期,以及关于进一步让渡主权的红利预期。其二,一体化事后预期存在功能性“溢出”效应。即,进一步扩大主权让渡,预期可充分实现上轮一体化的红利且带来新的红利。由于存在路径依赖问题,减少主权让渡逆一体化预期将带来巨大的损失。情形(2),国民主权让渡的红利预期下降。如图4中所示,主权让渡的预期红利曲线 $E(T)$ 由 $E_1$ 向内(靠近纵轴)移至 $E_3$ 。类似地,有两种情形出现会导致一体化红利预期下降。其一,上轮一体化事后实现的红利小于事先预期的红利。在此情形下,参与方基于适应性预期规则,将下调其给定主权让渡不变情形下的未来红利预期,以及关于进一步让渡主权的红利预期。其二,或者因路径依赖造成的预期损失很小,或者因新合作机会的预期收益很大,逆一体化预期净收益为正。譬如,东欧国家在冷战结束后选择退出华约,一方面是因为退出华约所带来的安全问题很小,另一方面是因为加入北约后可获得更好的安全保障。情形(3),国民主权让渡的红利预期不变。如果上述影响红利预期的因素都未出现,那么国民关于主权让渡红利的预期就不会发生改变。

第二类反馈,导致参与国国民关于国家间一体化的偏好发生变化,包括(4)(5)(6)3种情形。情形(4),国民一体化偏好分布向左倾斜。如图

2-1所示，国民偏好曲线由 AA 移动至 BB，由正态分布变为左偏分布。即，A 国国民自决意识下降，即使一体化红利预期不变，也愿意让渡更多的主权。偏差系数  $Cs_U$  下降幅度越大，则国民偏好分布向左倾斜幅度越大。新功能主义所描述的政治性“溢出”效应就属于这一反馈情形。一方面，两国国民互动联系增多，相互依存度上升，出现更多的共享文化和历史。另一方面，支持一体化的利益集团、超国家组织、跨国组织的可利用资源增多，可对国民偏好的形成带来更大影响。在此两种力量作用下，国民对新的共同体更加认同，更愿意将个人忠诚转移至新的共同体。情形（5），国民一体化偏好分布向右倾斜。如图 2-1 中所示，国民偏好曲线由 AA 移动至 CC，由正态分布变为右偏分布。即，A 国国民自决意识上升，即使一体化红利预期不变，也希望减少主权让渡。偏差系数  $Cs_U$  上升幅度越大，则国民偏好分布向右倾斜幅度越大。后功能主义一体化理论描述的政治性“回溢”属于这一反馈情形。其一，随着主权让渡对参与方的制约增多，参与方国民自决意识和国民身份认同增强，一体化问题日益成为国内政治焦点。其二，随着两国国民互动联系增多，不仅彼此之间文化、历史上的冲突被放大，还引发了新的矛盾。在此两种力量作用下，国民更强调国家身份认同，更不愿意将个人忠诚转移至新的共同体。情形（6），国民一体化偏好分布不变。如果上述影响国民偏好分布的因素都未出现，那么国民偏好曲线就不会变化移动。

### （三）第二轮博弈

当反馈效应超过临界点，新一轮博弈就会启动。所谓超过临界点，是指预期红利曲线或者政策无差异偏好曲线位置发生变化。上述两类 6 种反馈情形可形成 9 种反馈组合（见图 4），我们将对每个反馈组合下的国民博弈、国家博弈进行分析。不失一般性，我们假定一体化前后 B 国国民的身份认同、主权偏好和收益预期均没有任何改变，仅讨论 A 国变化对博弈均衡结果的影响。

第一种反馈组合，情形（1）（4）同时发生。博弈第一阶段，情形（4）发生，A 国国民偏好分布向左倾斜，由此导致政策无差异曲线由  $T_1$  向上推移至  $T_3$ ，如图 4-1 所示。情形（1）发生，给定红利分配比例，A 国预期红利曲线由  $E_1$  向外移到至  $E_2$ 。博弈第二阶段，给定 B 反应函数不变，A 国因政策无差异曲线上移、红利预期曲线外移，A 国反应曲线相应地由  $T_A^0(\alpha)$  上移至  $T_A^1(\alpha)$ ，如图 5-1 所示。由此，均衡点从  $E_0$  沿着 B 国反应曲线  $T_B(\alpha)$  向上移动至  $E_1$ 。即，情形（1）（4）反馈组合将导致均衡主权让渡由  $T_0^*$  增加至

$T_1$ , A 国均衡红利分享比例由  $\alpha_0^e$  下降至  $\alpha_1^e$ , 两国一体化得到深化和拓展, 总的一体化红利预期上升。

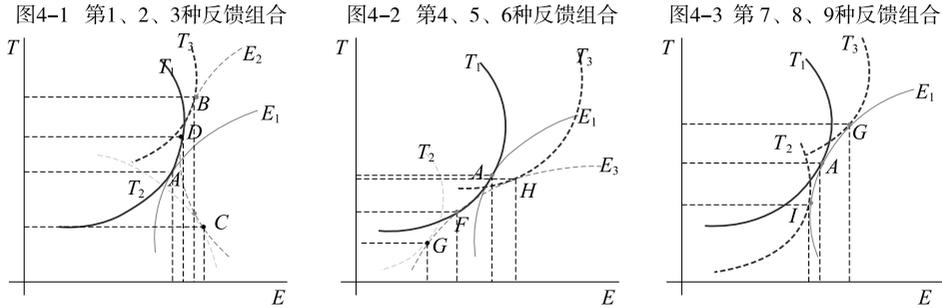


图 4 9 种反馈组合

资料来源: 笔者绘制。

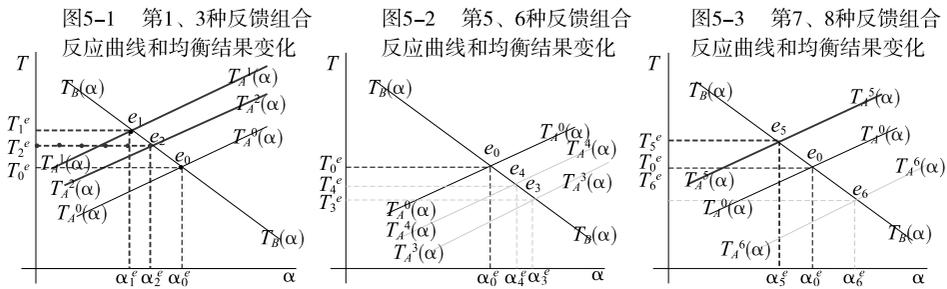


图 5 第二次博弈均衡

资料来源: 笔者绘制。

第二种反馈组合, 情形 (1) (5) 同时发生。博弈第一阶段, 情形 (5) 发生, A 国国民偏好分布向右倾斜, 由此导致政策无差异曲线由  $T_1$  向下推移至  $T_2$ 。情形 (1) 发生, 给定红利分配比例, A 国预期红利曲线由  $E_1$  向外移至  $E_2$ 。博弈第二阶段, A 国政策无差异曲线下移, 而红利预期曲线外移, 导致反应曲线移动方向不确定。因此, 情形 (1) (5) 反馈组合的均衡结果不能确定。

第三种反馈组合, 情形 (1) (6) 同时发生。博弈第一阶段, 情形 (6) 发生, A 国国民偏好分布没有变化, 政策无差异曲线保持在  $T_1$  位置。情形 (1) 发生, 给定红利分配比例, A 国预期红利曲线由  $E_1$  向外移至  $E_2$ 。博弈第二阶段, 给定 B 反应函数不变, A 国政策无差异曲线不变, 同时红利预期曲线外移, 故反应曲线相应地由  $T_A^0(\alpha)$  上移至  $T_A^2(\alpha)$ , 如图 5-1 所示。由此, 均衡点从  $E_0$  沿着 B 国反应曲线  $T_B(\alpha)$  向上移动至  $E_2$ 。即, 情形 (1)

(6) 反馈组合将导致均衡主权让渡由  $T_0^*$  增加至  $T_2^*$ ，A 国均衡红利分享比例由  $\alpha_0^e$  下降至  $\alpha_2^e$ ，两国一体化得到深化和拓展，总的一体化红利预期上升。

第四种反馈组合，情形 (2) (4) 同时发生。博弈第一阶段，情形 (4) 发生，A 国国民偏好分布向左倾斜，由此导致政策无差异曲线由  $T_1$  向上推移至  $T_3$ 。情形 (2) 发生，给定红利分配比例，A 国预期红利曲线由  $E_1$  向内移动至  $E_2$ 。博弈第二阶段，A 国因政策无差异曲线上移，同时红利预期曲线内移，导致反应曲线位置变化方向不能确定。因此，情形 (2) (4) 反馈组合的均衡结果不能确定。

第五种反馈组合，情形 (2) (5) 同时发生。博弈第一阶段，情形 (5) 发生，A 国国民偏好分布向右倾斜，由此导致政策无差异曲线由  $T_1$  向下推移至  $T_2$ 。情形 (2) 发生，给定红利分配比例，A 国预期红利曲线由  $E_1$  向内移动至  $E_3$ 。博弈第二阶段，给定 B 反应函数不变，A 国政策无差异曲线向下移动，同时红利预期曲线向内移动，故反应曲线相应地由  $T_A^0(\alpha)$  下移至  $T_A^3(\alpha)$ 。由此，均衡点从  $E_0$  沿着 B 国反应曲线  $T_B(\alpha)$  向下移动至  $E_3$ 。即，情形 (2) (5) 反馈组合将导致均衡主权让渡由  $T_0^*$  下降至  $T_3^*$ ，A 国均衡红利分享比例由  $\alpha_0^e$  上升至  $\alpha_3^e$ ，两国一体化遭到浅化和收窄，总的一体化红利预期下降。

第六种反馈组合，情形 (2) (6) 同时发生。博弈第一阶段，情形 (6) 发生，A 国国民偏好分布没有变化，政策无差异曲线保持在  $T_1$  位置。情形 (2) 发生，给定红利分配比例，A 国预期红利曲线由  $E_1$  向内移动至  $E_3$ 。博弈第二阶段，给定 B 反应函数不变，A 国政策无差异曲线向下移动，同时红利预期曲线向内移动，故反应曲线相应地由  $T_A^0(\alpha)$  下移至  $T_A^3(\alpha)$ 。由此，均衡点从  $E_0$  沿着 B 国反应曲线  $T_B(\alpha)$  向下移动至  $E_4$ 。即，情形 (2) (6) 反馈组合将导致均衡主权让渡由  $T_0^*$  下降至  $T_4^*$ ，A 国均衡红利分享比例由  $\alpha_0^e$  上升至  $\alpha_4^e$ ，两国一体化遭到浅化和收窄，总的一体化红利预期下降。

第七种反馈组合，情形 (3) (4) 同时发生。博弈第一阶段，情形 (4) 发生，A 国国民偏好分布向左倾斜，由此导致政策无差异曲线由  $T_1$  向上推移至  $T_3$ 。情形 (3) 发生，A 国主权让渡红利预期曲线保持在  $E_1$  位置不变。博弈第二阶段，给定 B 反应函数不变，A 国政策无差异曲线向上移动，同时红利预期曲线不变，故反应曲线相应地由  $T_A^0(\alpha)$  上移至  $T_A^3(\alpha)$ 。由此，均衡点从  $e_0$  沿着 B 国反应曲线  $T_B(\alpha)$  向上移动至  $e_5$ 。即，情形 (3) (4) 反馈组合

将导致均衡主权让渡由  $T_0^*$  增加至  $T_5^*$ ，A 国均衡红利分享比例由  $\alpha_0^e$  下降至  $\alpha_5^e$ ，两国一体化得到深化和拓展，总的一体化红利预期上升。

第八种反馈组合，情形（3）（5）发生。博弈第一阶段，情形（5）发生，A 国国民偏好分布向右倾斜，由此导致政策无差异曲线由  $T_1$  向下推移至  $T_2$ 。情形（3）发生，A 国主权让渡红利预期曲线保持在  $E_1$  位置不变。博弈第二阶段，给定 B 反应函数不变，A 国政策无差异曲线向下移动，同时红利预期曲线不变，故反应曲线相应地由  $T_A^0(\alpha)$  下移至  $T_A^6(\alpha)$ 。由此，均衡点从  $e_0$  沿着 B 国反应曲线  $T_B(\alpha)$  向下移动至  $e_6$ 。即，情形（3）（5）反馈组合将导致均衡主权让渡由  $T_0^*$  下降至  $T_6^*$ ，A 国均衡红利分享比例由  $\alpha_0^e$  上升至  $\alpha_6^e$ ，两国一体化遭到浅化和收窄，总的一体化红利预期下降。

第九种反馈组合，情形（3）（6）同时发生。博弈第一阶段，情形（6）发生，A 国国民偏好分布没有变化，政策无差异曲线保持在  $T_1$  位置。情形（3）发生，A 国主权让渡红利预期曲线保持在  $E_1$  位置不变。显然，在情形（3）（6）反馈组合下，新一轮博弈将不会发生，原有均衡将继续保持。

#### （四）稳态均衡

如果第一种、第七种反馈组合出现，第二轮博弈结束后一体化将得以深化和拓展（演进）。如果第二轮博弈结束后，第一种、第七种反馈组合继续重复出现，导致国民偏好持续向左倾斜，只要红利预期不下降，就会推动主权让渡持续深化和拓展。给定两国情形完全对称，在此情形下，主权让渡深化拓展与国民偏好向左倾斜将一轮一轮地彼此强化，最终结果是两国国民偏好彻底左偏，即  $P(U_1) = 1$ ，所有主权都将被让渡从而形成新的国家。这一稳态变迁过程就是新功能主义所推崇的一体化动态演进过程：一体化进程一旦启动，就会在溢出效应和路径依赖两大机制作用下自动进行，最终实现完全一体化。从煤钢共同体到欧盟属于此类稳态均衡调整过程的典型案例。

如果第五种、第八种反馈组合出现，第二轮博弈结束后一体化将遭到浅化和收窄（退化）。如果第二轮博弈结束的反馈过程中，第五种、第八种反馈组合继续重复出现，导致国民偏好持续向右倾斜，只要红利预期不提高，就会推动主权让渡持续浅化和收窄。给定两国情形完全对称，在此情形下，主权让渡浅化收窄与国民偏好向右倾斜将一轮一轮地彼此强化，最终结果是两国国民偏好彻底右偏，即  $P(U_0) = 1$ ，所有的一体化形态都会被完全解构，国家之间不再存在结构性合作安排。这一动态收敛过程，与后功能主义一体化理论描述的自我挫败、自我限制的逆一体化过程相一致。独联体就是逆一

体化过程的典型案例<sup>①</sup>。

出现第三种、第六种、第九种反馈组合，意味着国民偏好没有发生变化（停滞）。其中，第三种、第六种反馈组合意味着事前红利预期与事后红利实现存在差异。在适应性预期规则下，第二轮博弈启动时红利预期将因此增加/减少，导致博弈均衡主权让渡增加/减少。如第二轮博弈结束后，此两种反馈组合反复出现，那么红利预期将逐步收敛于实际红利，博弈均衡将收敛于稳态均衡点。如果第一轮博弈过程中红利预期与事后红利实现完全相符，那么就会出现第九种反馈，第一轮博弈均衡结果就是稳态均衡点。在此三种反馈组合下，国家之间可实现长期稳定的某种一体化形态，与自由政府间主义的描述相一致。东盟、加勒比共同体、非盟等非欧地区一体化长期处于此类均衡状态，世贸组织、联合国等国际组织也属于此类稳态的典型案例。

#### （五）成员数量问题

国家间一体化的一般理论虽然只涉及双边博弈，但同样可以分析成员数量扩大的问题。下面，我们将简单分析 A 和 B 两国就一体化达成一致后，国家 C 申请加入新共同体的情形。C 国加入一体化博弈后，参与方 A 作为理性谈判者，将不会单独考虑 C 和 B 的反应，而是将二者作为整体进行考量。对于 A 国而言，C 国参与将带来积极和消极两方面的影响。积极的一面，C 国参与导致共同体规模扩大，共同体多元性上升。这意味着，C 国参与既可因规模经济效应给 A 国带来一体化预期收益增加，又可能因多元性上升而提高 A 国国民对于共同体的认同感。给定其他因素不变，A 国一体化收益预期将因此增加，而 A 国国民偏好也将因对共同体认同感上升而向右偏移。消极的一面，C 国参与将改变 A 国与谈判对手的依存关系，并加大共同体的异质性。对于 A 国而言，C 国参与意味着谈判对手力量加强，自身谈判地位相对下降，从而可能导致红利分配份额下降。同时，共同体异质性上升，既可能造成 A 国国民对共同体的认同下降、一体化偏好左倾，又可能因弥合成本上升导致一体化净预期红利减少。

A 国到底该如何选择？答案将取决于 A 国对以上积极和消极两方面影响

---

<sup>①</sup> 苏联解体后，由前苏联共和国组成独立国家联合体。成立时，除波罗的海 3 国外，其他 12 个前苏联加盟共和国阿塞拜疆、亚美尼亚、白俄罗斯、格鲁吉亚（1993 年 12 月起）、吉尔吉斯斯坦、摩尔多瓦（1994 年 4 月起）、哈萨克斯坦、俄罗斯、乌兹别克斯坦、乌克兰、塔吉克斯坦和土库曼斯坦均为独联体正式成员国。2005 年 8 月，土库曼斯坦宣布退出独联体。2008 年 8 月 14 日，格鲁吉亚宣布退出独联体，并于 2009 年 8 月 18 日正式退出。2014 年 3 月，因为克里米亚独立入俄问题，乌克兰也正式启动退出程序。

的权衡。如果成员数量增加体现的多元化效应大于异质化效应，A国国民对规模扩大后的共同体认同感上升，那么其国民偏好将会向右倾斜，政策无差异曲线向右向下移动。如果成员增加带来的预期规模经济效益足够大，可以弥补A国可能出现的红利分享份额下降所造成的损失，那么A国面临的预期收益曲线将向上向左移动。在以上两种情况下，如果其他因素保持不变，则A国推动一体化深化、拓展和扩大的积极性越大。在此情形下，接受C国将共同体成员从A、B两国扩大至A、B、C三国，同时扩大共享的公共产品的范围、提高公共产品集体供应的比例，将是A国占优的理性选择；反之，A国基于理性考量，或者选择接受C国申请，或者选择缩小共享的公共产品的范围、降低公共产品集体供应的比例。

#### 四 结语

功能主义和政府间主义是国家间一体化理论的主流，两大学派之争贯穿着整个国际一体化实践的历史，至今未休。虽然两大阵营都承认对方有其合理性，但彼此根深蒂固的偏见使得理论整合从未得到应有的关注和尊重，很少有学者尝试将二者进行整合。坦率地讲，即使受到关注与重视，成功整合本身也非易事。一体化是一个全球现象而不仅是欧洲现象，但长期以来理论界似乎并未注意到这一点。无论新功能主义、后功能主义还是自由政府间主义都过于聚焦于欧洲一体化，忽略了对其他地区不同类型的一体化实践的研究，未能抽象出国家间一体化更基本的特征。本文旨在整合新功能主义、后功能主义一体化理论和自由政府间主义，构建关于国家间一体化的一般理论，不仅关注同一类型国家间一体化进程的静态和动态特征，也关注不同类型国家间一体化的动态和静态特征。

然而国家间一体化一般理论还有不少须进一步完善之处。首先，理论假设将国家间博弈简化为双边博弈，并对博弈双方进行了对称处理。这样分析起来固然比较方便，但失之于简单。其次，本文在分析国家意愿形成时，描述了民主国家投票选举的过程，但未考虑到其他方式产生的结果。虽然这并不影响模型的理论意义，但留出了进一步分析的空间。最后，本文虽然用博弈模型对国家间一体化进行了量化描述，但尚不能进行定量分析，模型的作用仍局限于指引分析的方向。未来可沿此理论路径，对国家间一体化的选择开展定量模拟，以不断完善理论的分析解释能力。

(责任编辑 王 帅)