

 中国社会科学院创新工程学术出版资助项目

公共参与人博弈与 社会最优契约

欧阳葵 王国成◎著

中国社会科学出版社



公共参与人博弈与 社会最优契约

欧阳葵 王国成◎著



图书在版编目 (CIP) 数据

公共参与人博弈与社会最优契约/欧阳葵, 王国成著. —
北京: 中国社会科学出版社, 2013. 5

ISBN 978 - 7 - 5161 - 2774 - 2

I. ①公… II. ①欧… ②王… III. ①社会契约—研究
IV. ①C973

中国版本图书馆 CIP 数据核字(2013)第 128924 号

出版人 赵剑英
选题策划 卢小生
责任编辑 卢小生
责任校对 张玉霞
责任印制 李 建

出 版 中国社会科学出版社
社 址 北京鼓楼西大街甲 158 号 (邮编 100720)
网 址 <http://www.csspw.cn>
中文域名: 中国社科网 010 - 64070619
发 行 部 010 - 84083635
门 市 部 010 - 84029450
经 销 新华书店及其他书店

印 刷 北京市大兴区新魏印刷厂
装 订 廊坊市广阳区广增装订厂
版 次 2013 年 5 月第 1 版
印 次 2013 年 5 月第 1 次印刷

开 本 710 × 1000 1/16
印 张 14
插 页 2
字 数 232 千字
定 价 40.00 元

凡购买中国社会科学出版社图书, 如有质量问题请与本社发行部联系调换

电话: 010 - 64009791

版权所有 侵权必究

摘 要

本书试图在传统的非合作博弈模型中引入一个公共参与人，其目标函数即为社会福利函数，并在纳什均衡基础上提出了社会最优契约的解概念，以期探索个体与集体之间复杂的行为关系，以及对中国经济问题的深入分析开辟一条新途径。

为了确定合理的社会福利函数，本书探讨了个体偏好与社会偏好、个体效用与社会福利之间的关系，利用偏好函数对集体偏好的各种通常定义和性质进行了改写。无差异准则和正反应准则是多数原则的充分必要条件；在集体理性假设下，无差异准则与帕累托准则是不兼容的。本书提出并证明了否决权定理，著名的寡头定理和阿罗不可能定理可视为这一定理的推论，也给出了帕累托扩展准则、等级制度和波达规则的一系列新的公理化条件。

在社会福利泛函框架下，我们提出了最低正义准则，并表明在标准福利主义下，唯一满足弱帕累托准则、匿名性、比率不可比性和最低正义准则的社会福利函数是纳什社会福利函数，唯一满足弱帕累托准则、匿名性、基数可比性（或序数可比性）和最低正义准则的社会福利函数是罗尔斯主义社会福利函数。

本书分别在策略型博弈和扩展型博弈中探讨了公共参与人和社会最优契约的概念。在完全信息策略型博弈中，我们在纳什均衡集上定义了社会最优契约，包括功利主义、罗尔斯主义和纳什最优契约，并证明了纳什均衡集是非空的紧集，从而由公共参与人的效用函数的连续性保证了社会最优契约的存在性。此外，不难看出，功利主义最优契约同时满足线性和基数单位可比性，罗尔斯主义最优契约满足基数可比性，而纳什最优契约则满足基数不可比性。

在不完全信息策略型博弈中，我们在贝叶斯（纳什）均衡的基础上定义了社会最优贝叶斯契约。在奥曼模型中，我们对知识函数的定义进行

了推广，并表明如果一个博弈的解是共同知识，那么该博弈的解是贝叶斯纳什均衡当且仅当其满足贝叶斯理性。经过适当的定义，我们也可以证明社会最优贝叶斯契约的存在性。

在完美信息扩展型博弈中，本书在子博弈完美均衡集上定义了动态最优契约，并证明了子博弈完美均衡集是紧集，从而可以证明动态最优契约的存在性。在不完美信息扩展型博弈中，本书分别在序贯均衡和颤抖手完美均衡的基础上定义了序贯最优契约和颤抖手完美最优契约。研究结论也证明，序贯均衡集和颤抖手完美均衡集都是非空的紧集，从而由公共参与人的效用函数的连续性可以保证序贯最优契约和颤抖手完美最优契约的存在性。

最后，本书考察了策略型博弈的相关均衡，提出了社会最优相关契约的概念，并证明了其存在性，同时也表明了社会最优相关契约的线性和不变性特征与社会最优契约的性质类似。

关键词：社会选择与社会福利 公共参与人 非合作博弈 纳什均衡 社会最优契约

ABSTRACT

This book attempts to introduce a public player, whose utility function is just the social welfare function, into the traditional noncooperative games and propose the definition of social optimal agreement as the solution concept based on Nash equilibrium.

In order to find an appropriate social welfare function, we explore many possibilities and impossibilities of aggregating individual preferences and utilities into social preference and utility. The preference function is defined, and the usual definitions and properties of collective preference are redescribed. The indifference principle and positive responsiveness together are sufficient and necessary conditions of the majority rule. Collective rationality, indifference principle, and Pareto principle are inconsistent. The veto power theorem is proposed and proved in this paper, and the famous oligarchy theorem and Arrow's impossibility theorem can be considered as direct corollaries. New characterizations of Pareto - Extension rule, hierarchy, and Borda rule are also given.

The minimal justice is proposed as a property of social welfare functional. Under formal welfarism, Nash's social welfare function is the only social welfare function satisfying weak Pareto principle, anonymity, ratio - scale non-comparability, and minimal justice; Rawlsian social welfare function is the only social welfare function satisfying weak Pareto principle, anonymity, cardinal full comparability (or ordinal level comparability), and minimal justice.

The concepts of public player and social optimal agreement are respectively explored in strategic and extensive games. In strategic games with complete information, the definitions of social optimal agreements, including utilitarian, Rawlsian, and Nash, are based on the set of Nash equilibria. The existence of social optimal agreement is established on the nonempty and compactness of

Nash equilibrium and the continuity of the public player's utility function. In addition, it is not hard to show that the utilitarian optimal agreements satisfy linearity and cardinal unit comparability, Rawlsian optimal agreements satisfy full comparability, and Nash optimal agreements satisfy cardinal noncomparability.

In strategic games with incomplete information, the definition of social optimal Bayesian agreement is based on Bayesian (Nash) equilibrium. In Aumann's model of incomplete information, the knowledge function is generalized, and it can be shown that if the solution concept of a game is common knowledge, then the game solution is Bayesian Nash equilibrium if and only if it satisfies Bayesian rationality. It can also be shown that the social optimal Bayesian agreements always exist if properly defined.

In extensive games with perfect information, the dynamic optimal agreements are defined upon the set of subgame perfect equilibria, which is nonempty and compact and thus guarantees the existence of dynamic optimal agreements. In extensive games with imperfect information, sequential optimal agreement and trembling hand optimal agreement are defined respectively upon the set of sequential equilibria and the set of trembling hand perfect equilibria. The nonemptiness and compactness of the sets of sequential and trembling hand perfect equilibria are also shown, and thus the existence of sequential and trembling hand optimal equilibria follows.

Finally, correlated equilibria of strategic games are also considered, and it is easy to show that the social optimal correlated agreements as defined always exist. The social optimal correlated agreement is similar in linearity and invariance to the social optimal agreement.

Key Words: Social Choice and Welfare, Public Player, Noncooperative Games, Nash Equilibrium, Social Optimal Agreement

前 言

摆在大家面前的这本学术专著，是我们在欧阳葵博士的学位论文基础上几经商改、最终由王国成定稿而成。

2009年9月，欧阳葵以优异的成绩考入中国社会科学院数量经济与技术经济研究所，在王国成的指导下攻读数量经济学专业的博士学位。起初，欧阳葵对王国成主持的一个关于收入差距演变的国家自然科学基金项目非常感兴趣，并希望能将有关研究作为博士学位论文的主题。但是，随着研究的进一步深入，我们越发认为，收入分配问题本质上是一个关于社会福利评价的问题。于是，我们逐步将研究兴趣聚焦到社会选择理论，而这本来就一直都很关注的一个领域。欧阳葵在博士一年级下学期的学习中，还选修了王国成作为研究生们开设的《博弈论》课程，并担任课程助理，这无疑促进师生一起系统而深入地学习了解、共同探讨研究博弈论的良好契机和条件，同时借助集合拓扑等现代数学知识、更进一步地提升对博弈论的兴趣和理解程度，也为分析研究收入分配等应用问题提供了更加有力的工具。

无论是对于社会选择理论还是对于博弈论，1950年无疑都是不平凡的一年。就在这一年，纳什发表了他的《 n 人博弈的均衡点》，创建了后来被称为“纳什均衡”的概念；阿罗则发表了他的《社会福利概念中的一个难题》，提出了后来被称为“阿罗不可能定理”的结论。前者标志着现代非合作博弈论的诞生，后者则标志着现代社会选择理论的诞生。到了今天，这两个学科的理论体系看起来已经非常完整，二者的结合也已经十分密切。如果还有人想继续在这两个学科中作出纯理论上的重要贡献，无疑是异常艰难的事情。然而，我们似乎都有喜欢挑战学术难题的个性和偏爱，王国成因材施教、因势利导地鼓励和引导生性倔强、发奋做理论研究的欧阳葵，在其攻读博士学位的研究生生涯中，几乎将全部精力都奉献给了博弈论和社会选择理论，并逐渐地发现和感悟到博弈论与社会选择理论

中的许多核心概念也许并不是那么完美，是有可能在传统的基础上提出一套相对成体系又富有新意的概念和方法。于是，在导师的果敢肯定和大力支持下，欧阳葵以超乎常人的毅力和吃苦精神完成了这篇博士论文。到目前为止，虽然还没有让我们完全满意的成果（特别地，每当欧阳葵博士想到纳什的那篇只有几页纸的博士论文，就会有巨大的压力），但是我们依然充满了信心，希望在今后的学术生涯中可以将博弈论继续发扬光大，也期待能让我国的社会科学界意识到博弈论的重要性。如果有可能的话，我们甚至还希望通过大家的努力，能让博弈论来统一社会科学的研究范式，因为我们有共同的感受和看法：我国社会科学界中之所以存在许多争论不休的问题，很大程度上就是因为许多概念和结论都是含混不清，缺乏统一的初始界定和逻辑起点，如果能将其进行严格的形式化，相信很多争论将不复存在。

人类社会的本质到底是合作还是非合作？这无疑是一个很难回答的问题，甚至很可能根本就是一个没有答案的问题。在冯·诺依曼与摩根斯坦的经典著作中，这个问题被转化为零和博弈与非零和博弈之分。但是，在纳什的处理下，无论是零和博弈还是非零和博弈，都可以统一用一种非合作的方式来处理，从而使得它们都拥有了一个共同的解概念——纳什均衡；至于博弈的合作性质，纳什的讨价还价模型则为之提供了一个非常漂亮的解概念，并且这个解概念也成为合作博弈研究方面的一个十分重要的基石。非合作博弈似乎在很长一段时间内成为博弈论研究与应用的主流，一些经典的博弈论教科书中甚至就干脆没有合作博弈方面的内容。然而，可以毫不夸张地说，合作博弈也是博弈论世界的半壁江山；合作博弈最无可争议的代表人物夏普利于去年获得了诺贝尔经济学奖，大概就可以说明这一点。

当然，本书所遵循的主要研究方法仍然是非合作性质的：公共参与人博弈的社会最优契约这一解概念直接建立在纳什均衡等非合作博弈解概念的基础上。但是，这一解概念所要强调的重点却是人类行为的合作性质；公共参与人并不是在人类社会一开始就存在的，而是由各个理性的参与人在非合作博弈的过程中为了超越“互害”、追求“互利”的结局而建构出来的一种“集体理性”的代理人。任何合理的博弈解概念都应当以个体理性为前提，这一点无论怎样强调都不过分。事实上，集体理性不过是理性的个体为了避免“集体非理性”的结局而进行有效交流的可能性后果；

对于集体理性的日益上升的需求其实就是现代社会中个体理性逐步发展成熟的标志。总之，如何在非合作的基础上达成各种可能的合作，正是反映了人类理性与智慧的闪光，也是人类社会未来发展的期待和希望。

我们衷心地感谢中国社会科学院的哲学社会科学创新工程和中国社会科学出版社提供的宝贵机会，能让欧阳葵的博士论文得以提高、可以有更多的机会呈现在世人面前，从而可以与更多的读者进行交流，也有更多的机会接受批评与指正！衷心地感谢中国社会科学院数量经济与技术经济研究所的各位同人和师长给予的关心和提携，尤其是他们对于学术的尊崇和执著的榜样力量，给予了我们巨大和持久的激励！

欧阳葵博士真诚地感谢王国成老师的指导：王老师不但把其引入了博弈论的广阔天地，还提供了非常优越的研究环境，以及生活上非常关键而无以回报的帮助；得益于王老师渊博的知识、广阔的胸襟和非常富有前瞻性的视野，总是一次次将研究指向带入更为深刻的层面。在欧阳葵本人的学术生涯以及职业发展方面，王老师也总是不遗余力地帮助，为其创造了很多宝贵的机会，也给出了许多非常有价值的建议。当欧阳葵艰难地攀登学术高峰而曾倍感孤独之时，王老师几乎是唯一能够给予其理解和鼓励的人；能够得到王老师的指导，似乎是其命中注定之事！

欧阳葵博士还衷心地感谢西北大学为其提供了良好的研究与教学的环境，经管学院深厚的人文气息、悠久的学术传统和自由宽松的学术气氛，为其进一步的研究提供了一片非常广阔而舒适的空间；衷心地感谢其爱妻，虽然在许多重要的时刻，欧阳葵都不能够陪在妻子身边，由此感到非常愧疚，妻子的默默支持是其一生最为宝贵的财富；同时，也衷心地感谢其家人、同学和朋友，他们的支持一直是欧阳葵博士勇往直前的不竭动力！

我们谨以此书献给立志于博弈论理论研究的朋友们，
献给欧阳葵博士的女儿——美丽的茜茜！

王国成 欧阳葵

2013年4月16日

目 录

前言	1
第一章 导论	1
第一节 非合作博弈的解概念	1
第二节 合作博弈的解概念	15
第三节 公共参与人博弈与社会最优契约	22
第四节 结构安排及相关数学符号说明	35
第二章 个体偏好与社会选择	38
第一节 偏好与理性	38
第二节 社会偏好的基本性质	40
第三节 无差异准则与多数原则	46
第四节 否决权与帕累托扩展准则	53
第五节 阿罗不可能定理与等级制度	58
第六节 波达规则	62
第七节 小结	69
第三章 个体效用与社会福利	71
第一节 偏好与效用	71
第二节 社会福利泛函	72
第三节 基数与序数不可比性	74
第四节 强中性与伯格森—萨缪尔森社会福利函数	76
第五节 萨普斯无差异准则	79
第六节 功利主义社会福利函数	82

第七节	纳什社会福利函数	86
第八节	罗尔斯主义社会福利函数	93
第九节	小结	97
第四章	公共参与人博弈的纳什均衡与社会最优契约	98
第一节	公共参与人博弈	98
第二节	纳什均衡及其存在性	99
第三节	社会最优契约及其存在性	102
第四节	功利主义、罗尔斯主义与纳什最优契约	104
第五节	社会最优契约的线性	108
第六节	社会最优契约的不变性	112
第七节	小结	115
第五章	不完全信息下的公共参与人博弈	117
第一节	类型与信息	117
第二节	信息、知识与共同知识	119
第三节	类型、知识与共同知识	123
第四节	信念结构与哈萨尼教条	126
第五节	贝叶斯博弈与贝叶斯均衡	127
第六节	社会最优贝叶斯契约	132
第七节	小结	134
第六章	公共参与人博弈的动态最优契约	135
第一节	扩展型公共参与人博弈	135
第二节	完美信息博弈及其纳什均衡	143
第三节	完美信息博弈的子博弈完美均衡	147
第四节	完美信息博弈的动态最优契约	150
第五节	不完美信息下的贝叶斯公共参与人博弈	153
第六节	序贯均衡与序贯最优契约	156
第七节	颤抖手完美均衡与颤抖手完美最优契约	159
第八节	小结	163

第七章 相关均衡与社会最优相关契约·····	165
第一节 公共参与人博弈的相关均衡·····	165
第二节 贝叶斯理性与相关均衡·····	169
第三节 相关均衡与纳什均衡的关系及其存在性·····	170
第四节 社会最优相关契约及其存在性·····	173
第五节 社会最优相关契约的线性·····	175
第六节 社会最优相关契约的不变性·····	176
第七节 小结·····	180
第八章 结论·····	182
参考文献·····	188

Contents

Chapter 1 Introduction	1
Section 1 Solution Concepts of Non – cooperative Games	1
Section 2 Solution Concepts of Cooperative Games	15
Section 3 Public – Player Games and Social Optimal Agreements	22
Section 4 The Organization of the Book and Some Special Symbols	35
Chapter 2 Individual Preferences and Social Choice	38
Section 1 Preferences and Rationality	38
Section 2 Basic Properties of Social Preferences	40
Section 3 Indifference Principle and Majority Rule	46
Section 4 Veto – Power and Pareto Extension – Rule	53
Section 5 Arrow ’ s Impossibility Theorem and Hierarchy	58
Section 6 Borda Rule	62
Section 7 A Brief Summary	69
Chapter 3 Individual Utilities and Social Welfare	71
Section 1 Preferences and Utilities	71
Section 2 Social Welfare Functional	72
Section 3 Cardinal and Ordinal Non – comparability	74
Section 4 Strong Neutrality and Bergson – Samuelson Social Welfare Function	76
Section 5 Suppes Indifference Principle	79
Section 6 Utilitarianism	82
Section 7 Nash Social Welfare Function	86

Section 8	Rawlsianism	93
Section 9	A Brief Summary	97
Chapter 4	Nash Equilibria and Social Optimal Agreements of	
	Public – Player Games	98
Section 1	Public – Player Games	98
Section 2	Nash Equilibria and Their Existence	99
Section 3	Social Optimal Agreements and Their Existence	102
Section 4	Utilitarian, Rawlsian, and Nash Optimal Agreements	104
Section 5	Linearity of Social Optimal Agreements	108
Section 6	Invariance of Social Optimal Agreements	112
Section 7	A Brief Summary	115
Chapter 5	Public – Player Games with Incomplete Information	117
Section 1	Types and Information	117
Section 2	Information, Knowledge, and Common Knowledge	119
Section 3	Types, Knowledge, and Common Knowledge	123
Section 4	Belief Structure and Harsanyi’s Doctrine	126
Section 5	Bayesian Games and Bayesian Equilibria	127
Section 6	Social Optimal Bayesian Agreements	132
Section 7	A Brief Summary	134
Chapter 6	Dynamic Social Optimal Agreements of Public – Player	
	Games	135
Section 1	Extensive Public – Player Games	135
Section 2	Nash Equilibria of Public – Player Games with Perfect Information	143
Section 3	Sub – game Perfect Equilibria of Public – Player Games with Perfect Information	147
Section 4	Dynamic Optimal Agreements of Games with Perfect Information	150

Section 5	Bayesian Public – Player Games with Imperfect Information	153
Section 6	Sequential Equilibria and Sequential Optimal Agreements	156
Section 7	Trembling Hand Perfect Equilibria and Trembling Hand Optimal Agreements	159
Section 8	A Brief Summary	163
Chapter 7	Correlated Equilibria and Social Optimal Correlated Agreements	165
Section 1	Correlated Equilibria of Public – Player Games	165
Section 2	Bayesian Rationality and Correlated Equilibria	169
Section 3	Correlated Equilibria, Nash Equilibria, and Their Existence	170
Section 4	Social Optimal Correlated Agreements and Their Existence	173
Section 5	Linearity of Social Optimal Correlated Agreements	175
Section 6	Invariance of Social Optimal Correlated Agreements	176
Section 7	A Brief Summary	180
Chapter 8	Conclusions	182
References	188

第一章 导论

合作博弈与非合作博弈是现代博弈论中两个不同的分析框架。我们的研究试图建立一个公共参与人博弈模型，即在传统的非合作博弈中引入一个公共参与人，其目标函数即为社会福利函数。在传统的非合作解概念基础上，我们提出社会最优契约的概念，并建议以其作为公共参与人博弈的标准解概念，从而可以非合作的方式实现某些合作的结果。在本章中，我们首先对非合作博弈和合作博弈的主要理论框架进行综述，然后在此基础上阐述公共参与人博弈的基本思想，并简要地说明以社会最优契约作为公共参与人博弈的标准解概念的合理性。^①

第一节 非合作博弈的解概念

1944年，冯·诺依曼与摩根斯坦合著的《博弈论与经济行为》一书出版，一般认为这标志着博弈论开始成为一门独立的学科。该书构建了一个较为完整的理论逻辑框架：将所有的 n 人非零和博弈转化为 $n+1$ 人零和博弈，而所有的 $n+1$ 人零和博弈都可以二人零和博弈为基础；而早在1928年，冯·诺依曼所证明的那个著名的极小极大值定理就为二人零和

^① 本章的综述可以为那些具备一定博弈论入门知识（例如，Gibbons, 1992; Rasmuson, 2006; 张维迎, 1996; 谢识予, 1997; 等等）的读者提供一个简要的系统性梳理。对于博弈论（例如，Myerson, 1991; Osborne 和 Rubinstein, 1994; 黄涛, 2004; 范如国, 2011; 等等）较为熟悉的读者可直接从本章第三节开始阅读。比本章更为全面的关于国内外博弈论理论研究历史与现状的综述可参见尚宇红（2003）、王国成（2009）、欧阳葵和王国成（2011），等等。原则上，正常阅读本书只需具备一定的数学分析基础和一点基本的点集拓扑知识即可。本书尽量做到“自容”（Self-contained），以保证具备一定数学基础但从来没有深入研究过博弈论和社会选择理论的读者也能够完全读懂本书，因此基本上所有在正文中严格给出的定理都将直接给出详细的证明（从而对于某些读者来说这可能显得过于啰嗦了），除了两个不动点定理之外。对于不动点定理的证明感兴趣的读者可参见博德（Border, 1985）。