“十二五”国家重点图书
经济学术译丛 · 当代制度分析前沿系列

博弈论与社会契约（第2卷·下）公正博弈

GAME THEORY AND THE SOCIAL CONTRACT (VOLUME 2)

—JUST PLAYING

肯·宾默尔
(Ken Binmore) 著

潘春阳
陈雅静 译
陈 琳

上海财经大学出版社

“十二五”国家重点图书

经济学术译丛
当代制度分析前沿系列

博弈论与社会契约

（第2卷·下）

公正博弈

肯·宾默尔
(Ken Binmore)
著

潘春阳
陈雅静
陈　琳
译

上海财经大学出版社

本书由上海文化发展基金会图书出版专项基金资助出版

图书在版编目(CIP)数据

博弈论与社会契约(第2卷)·公正博弈/(英)宾默尔(Binmore, K.)著;潘春阳等译.—上海:上海财经大学出版社,2016.1
(经济学术译丛·当代制度分析前沿系列)
书名原文:Game Theory and the Social Contract Ⅱ:Just Playing
ISBN 978-7-5642-1875-1/F·1875

Ⅰ.①博… Ⅱ.①宾… ②潘… Ⅲ.①博弈论—研究 ②社会契约—研究 Ⅳ.①F224.32 ②F246

中国版本图书馆CIP数据核字(2014)第073362号

□责任编辑 温 涌
□封面设计 周卫民

BOYILUN YU SHEHUI QIYUE

博弈论与社会契约

(第2卷)

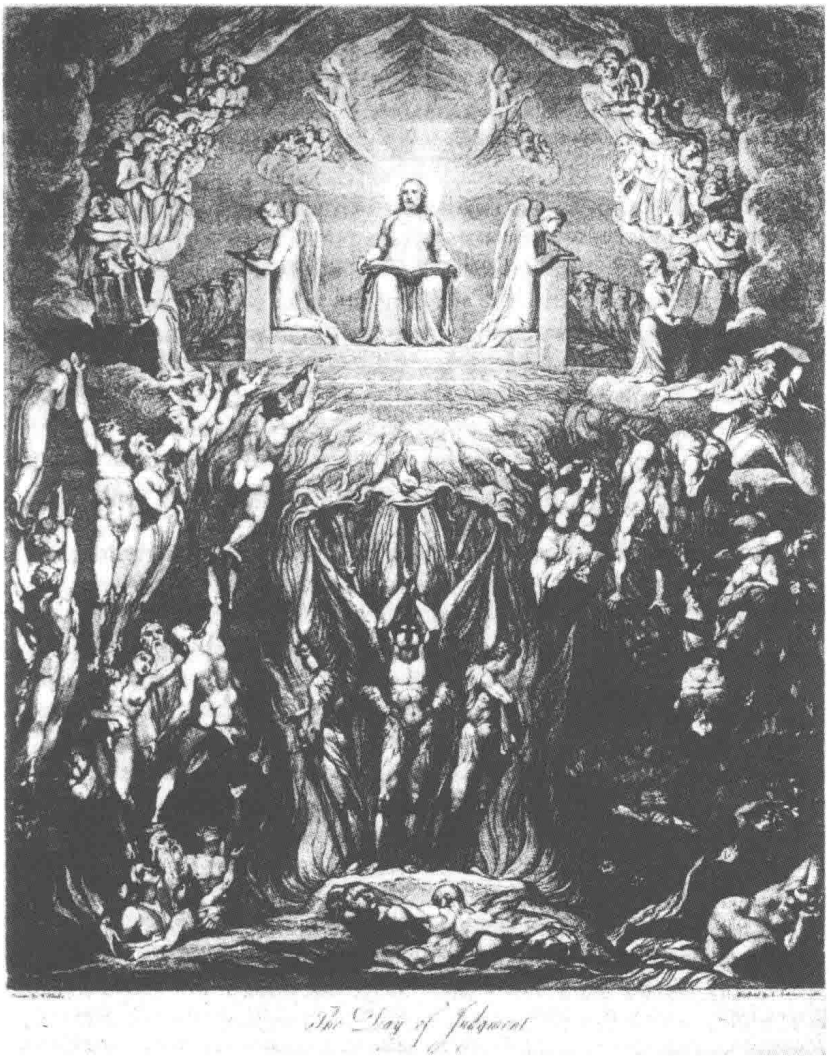
公正博弈

肯·宾默尔
(Ken Binmore) 著
潘春阳 陈雅静 陈 琳 译

上海财经大学出版社出版发行
(上海市武东路321号乙 邮编200434)
网 址:http://www.sufep.com
电子邮箱:webmaster@sufep.com
全国新华书店经销
上海叶大印务发展有限公司印刷装订
2016年1月第1版 2016年1月第1次印刷

890mm×1240mm 1/32 22.375印张 621千字
印数:0 001—3 000 定价:80.00元(上下册)

目　录



那些决定我们社会地位的人的语焉不详的反对，也经常足以使我们维系在由一个次优社会契约所决定的均衡路径之上。如果我们可以劝导人们更多地关注生存博弈在人死后将继续进行这一思想，而不仅是获得人们的口头支持，那么威廉·布莱克所面临的惩罚就能使最优的社会契约在当下实现。

第 3 章

理性互惠

让我总是做这个，我就会总是让你做这个，你就会让别人总是做这个。你在一个商业国家中承担了人的全部责任。

——潘克斯先生(Mr. Pancks)，引自狄更斯(Dickens)，《小杜丽》(*Little Dorrit*)

3.1 相互帮助

前一章以亚当和夏娃被驱逐出伊甸园而告终。在本书的剩余部分，再也不存在哲学王或者其他外部执行机构以维系某个他们一起制定的契约。唯一存在的“执法者”将是亚当和夏娃自己。

如果他们的生存博弈是如图 3.1(a)所示的单次囚徒困境，那么结局将变得十分悲惨。在存在一个哲学王以实施协议的情况下，在如图 3.1(b)阴影部分所示的合作收益区域内的任何一个收益对，都可以成为一个社会契约(参考第 1.2.1 节)。在不存在哲学王的情况下，亚当和夏娃将被局限于运行他们生存博弈中的均衡状态。但是，单次囚徒困境博弈只有一个均衡点——在这个均衡点中，每个局中人都会采用“鹰”这一强占

优策略并获得零回报。

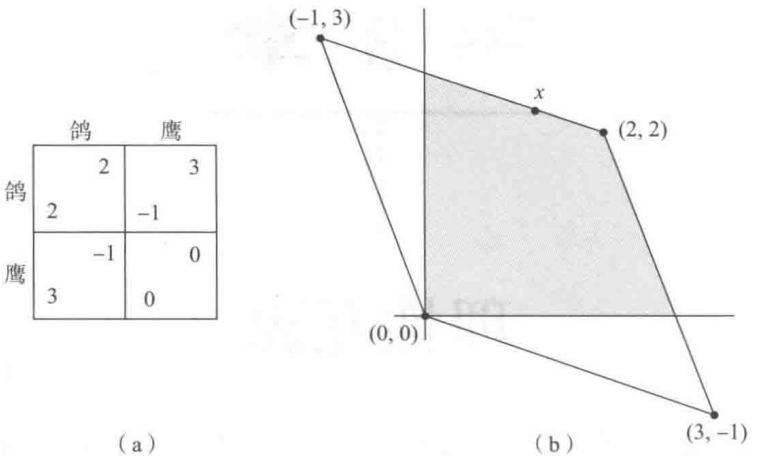


图 3.1　囚徒困境

然而,正如第1卷所反复强调的,遵循某些政治哲学家们的做法是错误的,他们致力于将我们的生存博弈构建为囚徒困境——或者是斗鸡博弈、性别战,抑或是其他单次博弈。如果我们的生存博弈的确是单次囚徒困境,那么我们将永远不可能进化为社会动物。尽管仍然包含了大量近于夸张的简化,但将生存博弈构建为一个无限期重复博弈显然更有意义。正如我在第1卷第2.2.6节所解释的,这样的建模方法解决了一个难题,那就是如何找到"正确"的博弈理论框架以刻画人类困境。真正重要的是,存在某个博弈是重复的,而我们选择重复哪个博弈在很大程度上是次要的。

在一个无限期重复博弈中,如果局中人仍然选择单次博弈中的一个纳什均衡,那么这仍然是一个均衡状态。在无限期囚徒困境博弈中,如果亚当和夏娃总是计划采用鹰策略,那么这也是一个均衡状态。这一均衡状态通常与霍布斯[251]提出的残酷的自然状态是一致的,是唯一一个能够替代独裁社会契约的选择。当然,还存在很多其他的纳什均衡点。在图3.1(b)阴影区域中的每一个收益对都代表了无限期重复囚徒困境博弈的一个纳什均衡结局(第1卷第2.2.6节)。特别地,亚当和夏娃可以

选择任何帕累托有效的收益对作为他们的社会契约，且这一收益对相比霍布斯自然状态是一种帕累托改进。

正如我在第1卷第2.2.6节所解释的，在一个无限期重复博弈中，支撑这种合作的纳什均衡的机制正是互惠性(reciprocity)——艾克斯罗德(Axelrod)[31]提出的著名的“针锋相对”(TIT-FOR-TAT)策略便是一个良好的例证。

合作的演化 这一策略要求一个局中人一开始便采用鸽策略，之后总是选择对手上一期采用的策略(第1卷第2.2.6节)。两个“针锋相对”的策略便构成了无限期囚徒困境博弈的一个纳什均衡。由于无论是背叛还是合作都是相互的，因而对“针锋相对”策略的最优反应，便是在所有时期都选择合作(假设贴现率并不是特别高)。因为采用“针锋相对”策略的局中人永远不会成为第一个背叛者，所以两个都选择了“针锋相对”策略的局中人总是能够保持合作状态。任何一个局中人都会对对手的策略选择做出最优反应。

人们认识到互惠性是人类社会性的主要动力，至少可以追溯到公元前7世纪，[1]但是，关于其内在机制可以辨识的现代表述，则不得不等待大卫·休谟的出现。正如休谟[267, p. 521]所解释的：

> 我学会了对别人进行服务，虽然我并没有感怀他的任何恩德，因为我可以预料，他会报答我的服务，并期望得到同样的另一次服务，并且也为了与我或与其他人维持同样的互助往来关系。而且相应地，在我为他服务之后，他会从我的行为中获得利益，他会被诱导以履行他的义务，因为他预见到他的拒绝可能产生的后果。

我怀疑，那些在50年代重新创立这一理念的博弈理论家们，可能完

[1] 施与那些施与别人的人，不施与那些不施与别人的人。——赫西奥德(Hesiod)[246]

全没有意识到休谟的思想。因为对他们而言,显而易见的是,单次博弈的应用范围十分有限,他们的工作仅仅是将纳什[381]当时新创造的均衡概念拓展至重复博弈。通常的情况便是,当某种思想出现的时机到来时,许多学者几乎同时想到了重复博弈理论的基本结论的不同版本[奥曼和马斯库勒(Aumann and Maschler)[28]]。之所以现在我们把这一结论称为无名氏定理(folk theorem),是因为没有人知道这一定理的发现应该归功于谁比较合适。该定理认为,单次博弈合作收益区域中的所有令人感兴趣的结局,也可以成为该博弈的无限期重复版本的均衡状态。因此,在重复博弈的情势下,我们并不需要一个哲学王来执行契约;也就是说,我们自己都做不了的事,外部执法机构也必将无能为力。

我不知道当时的博弈理论家们对这一发现的广泛意义的理解达到了什么程度,但是他们显然没有成功地将这一福音传播至生物群落,结果是,特里弗斯(Trivers)[524]在大约15年之后重新发现了这一基本思想,他成为一个更为成功的福音传道者。因此,无名氏定理背后的作用机制被认为是互惠利他主义(reciprocal altruism)的,尽管在这一机制运作时,并没有真正的利他主义包含其中。

在梅纳德·史密斯(Maynard Smith)[347]广受赞赏的《演化与博弈论》(*Evolution and the Theory of Games*),以及艾克斯罗德(Axelrod)[31]更为成功的《合作的演化》这两部名著出版十多年以后,严肃的博弈理论几乎仍然是演化生物学家和政治科学家们的未知领域。例如,杜格纳托夫(Dugnatov)[163, p. 11]的《动物间的合作》(*Cooperation among Animals*)告诉我们,冯·诺依曼和摩根斯坦对博弈论的贡献在于提出了囚徒困境!因此,我决定在本章详尽地讨论无名氏定理及其内涵,当然,首先我们有必要对这一定理在社会契约理论中的地位进行拓展回顾。

自利 自由主义者想当然地认为,那些使各种大家族团结起来的相互尊重和兄弟情义的共同纽带,也必然提供了防止大型社会分崩离析的黏合剂。事实上,对许多学者来说,自利动机能够提供足够的向心力以维

持社会运转的思想令人十分厌恶。我同意利他主义者相比利己主义者更能够成为合得来的伙伴,但我认为,人们需要成为利他主义者,才能使社会运转这一信念与人类历史的事实并不一致。陀思妥耶夫斯基(Dostoyevsky)[162, p. 35]在其自传《死屋手记》(*House of the Dead*)〔1〕中描述他在俄国沙皇集中营的囚徒经历时,非常清晰地阐明了这一点:

> 那里的绝大多数人都堕落得无可救药。诽谤和流言从未停息;这就是一个地狱、一个灵魂的黑夜。但是,没有人敢于反抗这一内在规律,而都选择接受监狱的规则;每一个人都屈服于它们。当做得过头的人来到监狱时,恐怖的气氛便笼罩了整个村镇。然而,新来的罪犯潜移默化地变得十分顺从,并融入了整体的环境。

总之,我们并不需要遵循埃尔斯特(Elster)[172]强调的爱与责任是连接社会的黏合剂的观点。更为现实的做法是,与詹姆斯·麦迪逊(James Madison)[215, p. 285]一起在大卫·休谟中寻找一个恰当的隐喻。〔2〕稳定的社会无须黏合剂以使其连接在一起,因为它们建立在与干石墙相同的基础之上,这些干石墙正是牧羊人用以为分清他们的财产而建造的。正如一只狒狒给另一只狒狒梳毛刷洗以期获得一次类似的服务作为回报那样,在干石墙中的每一块石头都被其邻居的重量固定在位,同时也贡献自己的力量将其邻居固定在位。

吉基尔博士(Dr. Jekylls)的社会相比海兹先生(Mr. Hydes)的社会无

〔1〕 道德哲学家们经常引用文学著作作为证据,以支持其所辩护的某种心理学理论的现象早已司空见惯。特别是荷马(Homer)和亨利·詹姆斯(Henry James)的作品似乎还广泛地成为灵感的来源。但确定无疑的是,人们从虚构作品中学习到的都是作者自身对人性的理解。然而,陀思妥耶夫斯基的《死屋手记》并不是一部虚构的作品,而是当时他亲眼所见并为之震惊的事实记录。

〔2〕 人类的幸福与繁荣……或许可以与拱顶建筑相类比,其中每一块石块,如果单独出来,都会掉落地面。但是通过各部分的相互支撑和组合,整个建筑结构便得以支撑。——大卫·休谟[266]