

PRISONER'S  
DILEMMA

# 囚徒的困境

冯·诺依曼、博弈论  
和原子弹之谜

[美] 威廉·庞德斯通 (William Poundstone) 著

吴鹤龄 译

中信出版集团

# PRISONER'S DILEMMA

## 囚徒的困境

冯·诺依曼、博弈论  
和原子弹之谜

[美] 威廉·庞德斯通 (William Poundstone) 著

吴鹤龄 译

## 图书在版编目 (CIP) 数据

囚徒的困境 / (美) 庞德斯通著, 吴鹤龄译. — 北京: 中信出版社, 2015.12 (2016.12 重印)  
书名原文: Prisoner's Dilemma  
ISBN 978-7-5086-5381-5

I. ①囚… II. ①庞… ②吴… III. ①博弈论—研究  
②冯·诺依曼, J. (1903-1957) —生平事迹 IV.  
①O225 ②K837.126.11

中国版本图书馆CIP数据核字 (2015) 第 177315 号

Prisoner's Dilemma by William Poundstone  
Copyright © 1992 by William Poundstone  
Chinese Simplified Characters copyright © 2015 by CITIC Press Corporation  
Published by arrangement with ICM Partners  
through Bardou-Chinese Media Agency  
ALL RIGHTS RESERVED.  
本书仅限中国大陆地区发行销售

## 囚徒的困境

著者: [美] 威廉·庞德斯通

译者: 吴鹤龄

策划推广: 中信出版社 (China CITIC Press)

出版发行: 中信出版集团股份有限公司

(北京市朝阳区惠新东街甲4号富盛大厦2座 邮编 100029)

(CITIC Publishing Group)

承印者: 北京楠萍印刷有限公司

开本: 880mm × 1230mm 1/32

版次: 2015年12月第1版

京权图字: 01-2015-5633

书号: ISBN 978-7-5086-5381-5 / K · 477

定价: 54.00元

印张: 11.5 字数: 310千字

印次: 2016年12月第12次印刷

广告经营许可证: 京朝工商广字第8087号

版权所有·侵权必究

凡购本社图书, 如有缺页、倒页、脱页, 由发行公司负责退换。

服务热线: 010-84849555 服务传真: 010-84849000

投稿邮箱: author@citicpub.com

"PRISONER'S"  
DILEMMA 目 录

## 1 二难推论

核武器的困境 // 006

约翰·冯·诺依曼 // 008

囚徒的困境 // 012

## 2 约翰·冯·诺依曼

神童 // 019

库恩统治时期的匈牙利 // 021

早期经历 // 023

普林斯顿高等研究所 // 025

克拉拉 // 027

个性 // 030

狂飙突进运动时期 // 038

世界上最聪明的大脑 // 043

## 3 博弈论

战争游戏 // 051

谁第一个发明了博弈论? // 055

博弈论和经济行为 // 057

分蛋糕 // 058

理性的游戏者 // 059

博弈树 // 060

博弈表 // 064

零和博弈 // 067

极小极大原理和蛋糕 // 069

混合策略 // 072

曲线球和致死基因 // 076

极小极大定理 // 079

$n$ 人博弈 // 080

## 4 原子弹

冯·诺依曼在洛斯阿拉莫斯 // 086

战时的博弈论 // 088

伯特兰·罗素 // 090

世界政府 // 092

比基尼岛的核试验 // 094

计算机 // 097

先发制人的战争 // 100

## 5 兰德公司

历史 // 109

敢于想入非非 // 115

语义学和芬兰语的语韵学 // 118

冯·诺依曼在兰德 // 121

约翰·纳什 // 122

马后炮 // 124

## 6 囚徒的困境

别克轿车的买卖 // 129

窃贼的信用 // 131

佛勒德-德莱歇实验 // 134

塔克的逸事 // 142

常识 // 148

文献中的“囚徒的困境” // 150

逃票的乘客 // 153

核竞赛 // 156

## 7 1950年

苏联的原子弹 // 163

“对战争不感兴趣”的人 // 166

尤里的演讲 // 167

福克斯事件 // 169

技术性突然袭击的性质 // 173

“为和平而侵略” // 176

弗朗西斯·马修斯其人 // 178

后果 // 181

公众的反应 // 184

难道这是一个试探气球？ // 188

麦克阿瑟的演讲 // 189

奥维尔·安德森 // 189

舆论界的反应 // 192

有多少颗原子弹？ // 194

尾声 // 198

## 8 博弈论及其不足

对博弈论的批评 // 205

效用和马基雅维利 // 208

人是理性的吗？ // 210

俄亥俄州的研究 // 212

## 9 冯·诺依曼的最后岁月

- 氢弹 // 221
- 一头猛虎 // 223
- 原子能委员会委员 // 225
- 带来希望的时刻 // 229
- 疾病缠身 // 232
- 巨星陨落 // 238

## 10 “胆小鬼”和古巴导弹危机

- “胆小鬼” // 245
- 志愿者的困境 // 250
- 志愿者困境实验 // 253
- 古巴导弹危机 // 254
- 疯子理论 // 263

## 11 其他社会难题

- 僵局 // 270
- 围捕牡鹿 // 271
- 不对称博弈 // 274
- 合作是正当的理由 // 276



霍华德的元博弈 // 280

逆向归纳悖论 // 282

## 12 最适者生存

稳定策略 // 290

基因中有背叛吗？ // 292

罗伯特·阿克塞尔罗德 // 295

一报还一报 // 299

一报还一报的麻烦 // 303

人工选择 // 307

镜中鱼 // 309

合作和文明 // 312

现实世界中的一报还一报 // 314

## 13 美元拍卖

逐步升级 // 320

苏比克的美元拍卖 // 323

现实生活中的美元拍卖 // 325

策略 // 330

理性的出价 // 332

在什么情况下博弈论不灵? // 334

最大数博弈 // 337

真空中的羽毛 // 342

**致谢 // 345**

**参考文献 // 347**

**译者后记 // 353**



PRISONER'S  
DILEMMA

---

1  
二难推论



这是一个著名的二难推论命题：有个人同他的妻子和母亲一起过河，中途在对岸突然出现一只长颈鹿，他立刻举枪向它瞄准。长颈鹿说：“如果你开枪，你母亲就没命；如果你不开枪，你妻子就完蛋。”这个人该怎么办呢？

这则经典的二难推论故事源于达荷美的波波族。<sup>①</sup>类似的稀奇古怪的故事、叫人难以做出决断的问题，在非洲民间传说中非常流行，其中许多故事还被西方的作家和哲学家借用过。在波波族的民间传说中，长颈鹿是会说人话的，而且说一不二、说到做到。用较为西方的方式和更加专业的术语，你可以像下面那样重述这则二难推论：你、你的配偶和你的母亲被几个发疯的科学家绑架，关在一个房间里，捆在椅子上动弹不得。房子里有一架古怪的机器，你正好可以触及其中一个按钮。一挺机关枪正对着你的配偶和母亲。墙上挂着一只钟，滴滴答答地走着，在阴森森的空气中发出令人恐怖的声音。一个科学家宣布，如果你按动面前的按钮，那么机关枪将瞄准你的母亲并把她击毙；如果你在 60 分钟内不去按按钮，那么机关枪将瞄准你的配偶开火。你仔细观察过这部残酷的机器，并且相信它会完成规定的程序。你该怎么办呢？

类似这样的二难推论有时会在大学的伦理学课程中进行讨论，当然也

---

<sup>①</sup> 达荷美是西非国家贝宁共和国的旧称。该国有 46 个主要部族，包括芳族、阿贾族、巴利巴族、约鲁巴族等。波波族是人数较少的一个部族。——译者注

不会有令人满意的答案。如果你坚持认为在这种情况下你只能什么也不做（不去按那个按钮，从而让机枪击毙你的配偶），理由是因为你什么也没有做，因此就没人能怪罪你。这显然是逃避责任的一种选择。你唯一可以选择的是确定你更爱你的配偶，还是更爱你的母亲，从而确定要保住哪一个人的性命。

在有些二难推论中，要让另外某个人同时也进行选择，这便使决策更加困难。在这种情况下，结果将取决于所有人做出的所有选择。在格雷戈里·斯托克（Gregory Stock）的《问题书》（*The Book of Questions*, 1987）中，有一则类似的但更具有挑战性的二难推论：“你和你深爱着的人分别被关在两个房间中，两人身边各有一个按钮，并且你们都知道，除非两人中有一人在规定的60分钟内按下按钮，否则两个人都要被处死；而先按按钮的人可以保住对方的性命，但自己将立刻被处死。你该怎么办呢？”

这里，有两个人要估量他们所处的困境，并独立地做出选择。不管哪个人去按动按钮都是生命攸关的。最棘手的问题在于：你应当在什么时候做出牺牲？这个二难推论强迫你在为自己还是为心爱的人提供一艘救生艇这个难题上做出抉择。

许多二难推论涉及某个人可能选择以牺牲自己为代价去保护另一个人，例如，父母可能为保护孩子而慷慨赴死，因为孩子显然有更长的生命之路。但不管应用什么准则（我们当然没有理由相信当事双方会应用同一个准则），有关救生艇的抉择都存在以下三种可能结局。

1. 当事双方在谁应当牺牲、谁应当活下去这个问题上达成共识时，前者（牺牲者）应该去按按钮以挽救后者。这种结局是最容易被接受的。

2. 双方都决定保护对方：母亲决定保护女儿，因为女儿会活得更加久一些；而女儿决定保护母亲，因为母亲给了她生命。在这种情况下，结局取决于谁抢先按下按钮。

3. 最令人不安的情况是双方都认为自己应该活下去。这样的话，没有人去按按钮，而时钟正在滴滴答答地走下去……

让我们来想象一下第三种情景：在计时开始之后已经过去了 59 分钟。你希望你所爱的人会去按按钮，但她（或他）却没有按（我们假定当另一方按了按钮后，幸存者会立即得到通知）。这时，你还有时间把各种可能性都仔细考虑一遍，因为有些人也许会琢磨整整一个小时才能决定谁应该活命，或者鼓起勇气去按按钮。但整整 59 分钟过去了，却没有任何动静，于是你就应该开始考虑你所爱的人是否已经决定由你来做出牺牲了。

发誓决不去按按钮是毫无意义的，即使在最后一秒钟。不管你怎样以自我为中心，你都没有能力逃过一劫。总有人要死，这是这个二难推论的必然结局。如果你所爱的人不愿意做出牺牲，那么你最好成全她（或他）。记住，你是真心爱着那个人的。

在理想情况下，在最后关头你会去按按钮的，而你所爱的人可能也想这么做。这就是你拖延着直到最后关头来临的全部理由。你想把在最后关头去按按钮的机会留给对方，但她（或他）却没有这么做，在这种情况下，也仅仅在这种情况下，你才会去按按钮。当然，你所爱的人可能也是这么打算的。

有两个因素使“双方都企图拖延到最后一秒”这种情况复杂化，那就是反应时间和时钟精度。那架该死的机器是不会有一丝一毫同情心的，一到时间它就会把你们两个立刻杀死。因此，你（或者是对方）必须抢在这种结局发生之前完成上述决策过程并迅速按下按钮。此外，墙上的挂钟并不一定与机器精确同步。当然疯子科学家会这样说，但因为他是疯子，说的未必可信。为安全起见，为了确保你是在规定时限内按了按钮，实际上你必须提前一点点。在采取“等待直至最后一刻”这种策略的情况下，这是最关键的一个决定！

你与也在拖延着时间的爱人的处境相同。如果双方都只在最后时刻来临时才下定决心，其后果将难以预料。其中一个人也许正好比另一人抢先

一步按下按钮，也可能双方都错过了时限而被处死。事实上，结果肯定是随机的，偶然性远大于合理性。

在古怪的房间里做出孤注一掷的决定这类问题在哲学论著中比比皆是，从而赢得了“问题盒”这样一个名称。这类二难推论问题为什么能引起人们如此的兴趣？部分是因为这种异乎寻常的困境使人感到新奇、刺激。但如果它们仅仅是一些测验智力的难题，并未与我们的个人经历产生共鸣，自然不会引起人们这么大兴趣的。

当然了，现实生活中的二难推论不是由发疯的科学家制造出来的，而是由于我们的个人利益同其他人的利益发生碰撞，或者是同社会利益发生冲突，从而以各种各样的形式建立起来的。我们天天都面临着艰难的选择；有时候，我们做出选择的方式同期望的方式很不一样。由二难推论引出的内在问题虽然简单，却令人十分困惑：在每一种情况下，是否都存在合情合理的行动方案呢？

## 核武器的困境

1949年8月，苏联在西伯利亚爆炸了它的第一颗原子弹，美国对原子能的垄断地位至此结束。世界上有了两个核大国这种局面，比西方观察家曾经预期的要早得多。<sup>①</sup>

苏联的原子弹激发了核武器竞赛，这种竞赛带来的某些后果是容易预见的——每个国家都想武装到这种程度，能发动一场快速压倒对手的核攻击。许多人意识到，这会导致令人无法接受的二难推论。在世界历史上第一次出现了这种可能性——某国期待通过一次闪电式的打击使敌国从地球上消失掉。在危机时期，按动核按钮的诱惑几乎是不可抗拒的。同样重要

---

<sup>①</sup> 许多西方观察家曾经预计苏联要到20世纪50年代中期甚至更晚才能掌握核技术。——译者注



的是，每个国家都害怕自己成为别国突然袭击的牺牲品。

20世纪50年代，在美国和西欧，曾有许多人主张美国对苏联发动一次直接的、不需要理由的核打击。人们为这种核打击赋予了一个委婉的名称，叫作“先发制人战争”。有这种想法的人认为，美国应该抓住时机，通过核讹诈或突然袭击来建立权威、统治世界。你也许认为，只有极端分子会支持这样一个计划。事实上，“先发制人战争”的运动在许多十分优秀的知识分子中也获得了支持，其中包括两位当代最出色的数学家：伯特兰·罗素<sup>①</sup>和约翰·冯·诺依曼。一般说来，数学家通常不是由于他们的政见或对世界的看法而闻名于世的；而且，罗素和冯·诺依曼是两个完全不同的人，但在“世界上不应该有两个核大国共存”这一点上，他们恰恰站到了一起。

罗素是“先发制人战争”这场运动的主要推动者，他热衷于宣传“苏联具有核摧毁能力”，除非苏联对由美国主导世界这种状态放弃主权。在1947年的一次演讲中，罗素说：“我倾向于认为俄罗斯人会默认美国主导世界这种状态：否则，世界将经历由此造成的战争，并出现一个独一无二的政府，因为这是世界的需要。”

冯·诺依曼的态度更加强硬，他热衷于出其不意地使用核武器先发制人。《生活》(Life)杂志曾经引用冯·诺依曼的言论：“如果你问为什么明天不用原子弹去轰炸他们，我倒要问为什么不今天就去轰炸呢？如果你说今天5点去轰炸，那么我要问为什么不是今天1点就去轰炸呢？”

他们两个人都对苏联没有任何好感。他们相信先发制人战争是逻辑的必然，是解决核扩散这一死结的唯一合理方案。在《新联邦》杂志(New Commonwealth)1948年1月号鼓吹先发制人战争的一篇文章中，罗素写道：“(对于先发制人战争)我已经提出的理由就像数学证明一样，是如此

---

<sup>①</sup> 伯特兰·罗素(Bertrand Russell, 1872—1970):英国数学家、哲学家。其数学巨著《数学原理》闻名于世、影响深远。他是至今唯一一位获诺贝尔文学奖的数学家。——译者注