

社会心理与应用系列

Cooperation in Social Dilemmas

The Power of Trust

社会困境中的合作 信任的力量

陈欣◎著



科学出版社

本书是全国教育科学规划教育部重点项目“信任对学生合作行为的影响机制研究”
(DBA130212)的最终成果

社会心理与应用系列

社会困境中的合作 信任的力量

Cooperation in Social Dilemmas
The Power of Trust

陈欣◎著

科学出版社
北京

内 容 简 介

社会困境是指个体在面对个人利益最大化（不合作）与群体利益最大化（合作）的冲突时个人的选择困境。信任是指个体在缺乏足够信息来判断他人动机、意图和行为的情况下，以对他人意图和行为的积极预期为基础，将拥有的资源交给他人处置并承担相应风险的意愿。本书通过文献整理和实证研究，论证和阐述了社会困境中个体之间的人际信任是走出社会困境的关键因素，培育信任、引导合作在社会困境中具有重要作用。

本书可供心理学科研工作者和心理学爱好者参阅。

图书在版编目（CIP）数据

社会困境中的合作：信任的力量 / 陈欣著. —北京：科学出版社，2019.9
ISBN 978-7-03-060216-9

I. ①社… II. ①陈… III. ①社会心理学 IV. ①C912.6-0

中国版本图书馆 CIP 数据核字（2018）第 292287 号

责任编辑：乔宇尚 崔文燕 南一荻 / 责任校对：王晓茜

责任印制：李 彤 / 封面设计：润一文化

编辑部电话：010-64033934

E-mail: edu_psy@mail.sciencep.com

科学出版社出版

北京东黄城根北街 16 号

邮政编码：100717

<http://www.sciencep.com>

北京虎彩文化传播有限公司 印刷

科学出版社发行 各地新华书店经销

*

2019 年 9 月第 一 版 开本：720×1000 B5

2019 年 9 月第一次印刷 印张：13 3/4

字数：247 000

定价：89.00 元

（如有印装质量问题，我社负责调换）



从“囚徒困境”（prisoners' dilemma）的故事中我们认识了社会困境，社会困境反映出的问题是人类会因自作聪明而作茧自缚，个人理性有时会导致集体的非理性。经典的“囚徒困境”被扩展成“重复的囚徒困境”，表明类似的困境会一而再地出现在人们的生活中，并且假如每个参与者都有机会去“惩罚”另一个参与者前一回合的不合作行为，当不同策略的参与者将这一行为重复了很长时间之后，从利己的角度可以判断，最终“贪婪”的策略趋向于减少，可能出现更多“利他”的策略。这个重复的博弈被进化心理学用来说明人类是如何通过自然选择使得一种利他行为的机制从最初自私的机制中进化而来的。

社会困境中的合作问题一直以来都有很多不确定因素，其作为一个开放式的研究范式吸引着很多学者。2006年，陈欣在南京师范大学读博士期间，也对这个选题感兴趣，并围绕这个选题做了一些研究工作，可以说，早期的研究作为《社会困境中的合作：信任的力量》的完成奠定了良好的基础。

在“囚徒困境”的策略竞赛中，“以牙还牙”被认为是一种好的策略，参与者以合作的方式出场，然后采取对手在前一回合用过的策略。然而，更好的策略是“宽恕地以牙还牙”，即对手出现了背叛行为，参与者仍然要以最小概率与其合作。这是考虑到偶尔出现的错误信息的传达，当出现错误信息时，“宽恕地以牙还牙”能给参与者一个从背叛中复原的机会，因此是最佳的。从这个最佳策略可以看到，要达成社会困境中的合作，一个重要的条件是“友善”，完全自私的策略是不会存在太长时间的。换一种思路，假如“囚徒困

境”中的两个人之间非常信任，如父与子，或者是很亲密的朋友，那么他们是很难出卖彼此的。进化心理学认为，人类的合作是利他行为进化的产物，从长远来看，利他、互惠、友谊、良好的声誉都可以让个体受益，是一种适应性行为。

在建立合作联盟的过程中，互相信任、避免欺骗无疑是重要的因素。可以说，社会困境中的合作将信任作为合作行为的前因变量，对于理解人们合作的心理机制有重要的作用。信任是一种复杂的心理状态，是个体因为预期到将来会出现的积极回报而做出的亲社会行为反应。同时，信任不仅是一种对回报的积极预期，还包含了个体与其他人及整个社会的关系，具有明显的关系取向。

如果你想和一个人交朋友，不要急于相信他，应该先看看他是否可靠，有些人只是贪图利益才与你结交，这样的关系绝不会持续到你有困难的时候。对手的“信任”，包含了比“友善”更丰富的内涵，人们之所以愿意合作，很多时候是相信对手不会过度地利用和剥削自己，这种信任可能是合作的心理学起源。在很多历史时期，首领或者国王利用“联姻”的方式来建立合作关系，因为在无法监管的情况下，为了确保氏族、部落或王国之间可以顺利合作，只能通过亲缘关系增加信任。在现代社会的社会文化和组织制度的各个层面，人们对信任也有着持续的关注。有人说“信任是社会系统的润滑剂”，也有人说“信任是预测经济成功的指标”。在现代企业制度下，劳动力自由流动程度较高，尤其是一些脑力劳动者，很多都无法被量化和监督，而员工在工作中的投入程度是效率和效益的重要影响因素，因此“信任”也具有社会困境的特点。

信任之所以变得越来越重要，是因为信任与合作有着密不可分的关联。信任与合作、互惠、风险三个概念常常同时出现，合作在很大程度上是建立在个人对群体其他成员期望的基础上的，也就是说相信其他成员也会采取相同的合作行为，信任对方不会利用自己的弱点来攻击自己。合作常常能得到互惠的结果。在信息交流不畅、信息不充分的条件下进行合作决策，互惠具有不确定性，这使得信任总是面临着一定的风险。应该说，人们对未来持续合作关系的期待维持着对他人行为的信任。信任反映了人与人之间关

系的某种心理契约，它能降低合作的成本，是合作关系的起点、前提和基础。

陈欣在该书中对社会困境、合作、信任进行了系统的阐述和分析，并且对于与信任、合作相关的一系列因素，如社会价值取向、人际关系、奖惩、声誉等，进行了理论梳理和实证研究，这对于未来的研究而言，有一定的参考价值。在此之前的社会困境研究，大多以期刊论文的形式存在，而该书坚持理论分析与实证研究相结合的思路，对近年来的文献进行了梳理和整合，把握该主题研究的发展脉络，阐述了社会困境的实质和核心，并对社会困境研究的现状进行了分析，在很多问题上提出了自己的见解。对于想要查阅和了解这方面内容的学者而言，该书提供了一个不同的视角来认识社会困境。

社会困境中的合作研究涉及心理学、经济学、概率论、社会学、教育学等多个学科，该书更多的是从心理学的视野来理解社会困境，试图引导人们更好地走出社会困境，运用家庭环境、亲子关系和心理发展阶段的特点来进一步培育信任、促进合作，这对于更好地理解人类社会的行为具有一定的指导作用。

在这个专题的研究上，陈欣一直走走停停，有进步也有停滞，《社会困境中的合作：信任的力量》的出版体现了作者自己的思考和探索，可以看成是一个阶段性成果。作为她的博士生导师，我希望她能在学术之路上坐得住冷板凳，坚守学术理想，积极思考、勤于实践、不断积累，创作不止。

同时，我也相信，随着人们对社会困境的研究越来越具体化，对一些有价值的研究结果的推介可以减少或者避免某些困境的产生，改善人们的社会生活环境，引导人们做出合作行为，营造一个更友善、可信任的社会环境。人与人之间、人与环境之间少一点精致的利己主义，多一些公共意识，共同开创更好的未来。

叶浩生

2019年5月3日于广州



现在冬日里很多的北方城市，常见的不是蓝天，而是雾霾。人们不停地抱怨，可又无能为力，大面积的空气和水污染问题成为当代人“会呼吸的痛”。草原逐渐沙化，近海海域几乎无鱼可以捕捞，地下水水位连年下降，生态环境的恶化不是突然发生的，在这个逐渐变化的过程中，每个人都可能是让环境变糟糕的推手。

在人类的发展史上，类似的情形并不鲜见，工业化过程中英国伦敦曾被称为“雾都”，美国的匹茨堡曾经下过“黑雨”，这些都是大自然被过度透支的结果。人类消耗的各种资源是对大自然的开发和利用，当使用的限度超过大自然的再生产和净化能力时，人类和大自然会同时沦为受害者。当个体不停地抱怨糟糕的环境，呼喊自己是环境的受害者时，其实自己也是这个结果的推手。对个人而言，这也许是微不足道的，但当人人都如此时，放大到群体范围，就会造成环境的大面积污染，个人也成为污染的受害者。

人与环境的关系如此，很多时候人与群体的关系也会出现同样的困境，如“滥竽充数”的成语故事，当个体都卖力地演奏时，演奏的效果很好，即使有一两个南郭先生不发出声音，也不会被发现；假如故事中的大多数人都不好好演奏，那么演奏的整体效果就会很糟糕，达不到理想的效果，整个乐队会被解散，乐师也会被辞退。个人蒙混不作为导致群体利益受损害，最后个人利益也同样受损害，学术界将这种现象命名为“社会困境”，也称社会两难（social dilemma）。

社会困境具有如下属性。

1) 无论社会组织中其他成员是否合作，个体不合作所得的收益要高于其合作的收益。

2) 行为决策的结果依赖于他人的行为选择。

3) 与合作行为相比，不合作总是对他人利益有害的。

4) 组织内成员通力合作要比他们都不合作时每个成员的获益更好。

因此，社会困境是指人们在面对个人利益最大化（不合作）与群体利益最大化（合作）这两个冲突时，一种左右两难的选择困境。对个体而言，保留自己的利益，通过“搭便车”（free-ride）行为享受其他组织成员的贡献比选择合作更有利可图。这种个体的获益受到集体贡献率的制约，因为当集体中的每个成员都选择合作时，集体利益实现最大化，这对个体利益的长远发展有着深远的意义。当集体中的个体都按照各自利益行事时，集体利益受损，个体在集体中的发展也会受限，个人利益也会因此受损。

在经典的“囚徒困境”中，当两个囚徒都觉得认罪是自己的最好选择时，等待他们的却是牢狱的惩罚，而不是无罪释放。也就是说，个人原以为的最好的选择却给所在的群体带来了不好的结果，这个结果更是个人不能够接受的。“理性人”得出一个悲观的结论：不想坐牢的两个囚徒一起坐牢。

在现实生活中，“理性人”很难存在。现实生活中人们表现出的理性非常有限，并且很少有人会以绝对自私的方式行动。因此，社会困境引发学者的兴趣主要基于两个原因：首先，越来越多的社会学者和行为学者认识到引发社会问题的部分原因是群体中个人的利己主义行为，人们希望自己能从选择中获利更多。其次，在社会困境中进行事实上的决策选择时，有些人做出不同于“理性人”假设的非理性决策，他们较少考虑个人利益，更多考虑的是群体利益，表现为一定程度上的利他和互惠。

近年来，人们对社会困境的研究在社会学、心理学和实验经济学等领域相互交叉，在数理博弈的基础上，采用行为实验法和认知神经科学等多元方法，取得了越来越丰富的研究成果。目前关于社会困境的研究，心理学方面以实证研究为主，虽然研究内容较为丰富，但还没有能够形成统一的理论框架和研究结论。

社会困境的实验研究表明，人们面对风险和不确定性的决策情景时，并不知道什么是最好的策略，更不知道自己该如何决策。实际上，个体的社会价值取向、对策策略、任务结构、奖惩条件、沟通、年龄、信息反馈方式（群体反馈、个人反馈）和榜样示范作用等都在不同程度上影响着社会困境中的合作。

目前的研究表明，结构性和动机性可以增加人们在公共物品困境中的合作概率。结构性解决方案关注的是社会困境中结构性参数对个人合作的影

响，结构性因素包括改变合作的收益矩阵参数和收益分配原则，以及引进奖惩系统等，推测追求个人利益最大化可能是人们不合作的原因。动机性解决方案主要关注个人处于社会困境中的心理状态对合作行为的影响，具体途径包括提升成员之间的信任、增强成员的集体认同感，或者使人们相信他们的贡献对集体的成功至关重要。从动机性因素来看，害怕和贪婪是人们选择不合作的原因。

在对结构性和动机性两类影响因素进行整合的过程中，越来越多的学者认为这两类因素在一定程度上是相互依赖的。结构性因素的调整会诱发动机或心理上的改变，而保证结构性调整的作用效果又必须以个人动机的改变为前提，有代表性的研究是在公共物品困境中引入惩罚制度。在梳理惩罚与合作的关系时，信任的作用崭露头角。

1) 探讨惩罚与合作的关系时，惩罚作为一种制裁方式，可以提高合作水平。但是，由于社会困境情境的广泛性和复杂性，惩罚具有较高的成本，操作起来具有一定的难度，惩罚很难成为长期有效的措施。信任作为一种内在的心理过程，在对惩罚与合作的研究中具有重要的作用，研究发现惩罚可能破坏了人们内在的合作动机——人际信任。

2) 信任是一种普遍存在的心理过程，人际信任往往能够影响社会困境中的合作，这种影响不仅体现在惩罚与合作的关系中，还体现在其他变量中，如社会价值取向、感恩、人际属性等。信任这个变量在社会困境中的合作研究中起着承上启下的作用，是合作的心理基础。

3) 信任博弈成为社会困境的一种研究范式，在信任博弈中，信任的测量通过合作水平的测量得以表现。

4) 信任与合作是密切联系但并不相同的两个概念。从心理学和教育学的视角来看，信任作为一种内在合作动机，具有内隐性和很强的可塑性。积极构建和培育人们之间的信任，可以引导人们走出社会困境，营造一个共赢的和谐社会。

本书从心理学的视角出发，对社会困境中的合作的影响因素进行梳理，发现在影响合作的结构性和动机性因素中，信任架起了一座沟通的桥梁。信任作为一种内在心理动机，能对合作的心理机制给予一定的解释。

信任与合作及其他变量之间存在一种理论上的关系，关系结构如图 0-1 所示。

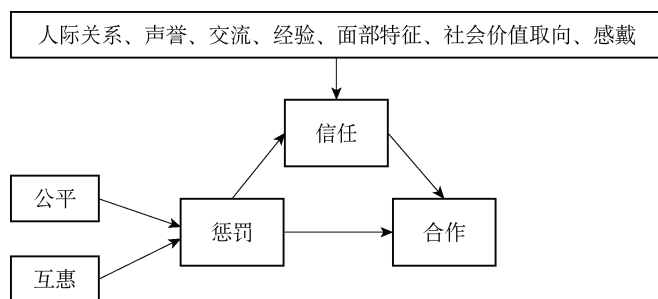


图 0-1 变量之间的关系结构图

本书还对培养信任、促进合作提出了心理学方面的途径和本土化思考，认为无论是身处家庭、学校还是社群组织，只要走出社会困境，人人都能有所作为。在此，我借用诗人雪莱的表达：“冬天来了，春天还会远吗？”即使一个人身处寒冷的冬天，只要心中有亮光，就能嗅到春天的气息；即使被困境所阻，只要心中有亮光，头顶的乌云总能被穿透。



序（叶浩生）

前言

第一章	社会困境的内涵和类型·····	001
	第一节 社会困境的内涵·····	003
	第二节 社会困境的类型·····	006
第二章	合作的前因变量和理论·····	013
	第一节 合作的前因变量·····	015
	第二节 合作的理论·····	021
第三章	社会困境及合作的研究现状与趋势·····	031
	第一节 合作的研究现状·····	033
	第二节 社会困境研究的趋势·····	036
第四章	信任的概念、操作和影响因素·····	043
	第一节 信任的概念·····	045
	第二节 信任的操作性定义·····	046

第三节	信任的影响因素	050
第五章	信任的类型、理论和生理基础	061
第一节	信任的类型	063
第二节	信任的理论	069
第三节	信任的生理基础	076
第六章	信任影响合作的心理机制	083
第一节	信任与合作的关系辨析	085
第二节	信任促进合作的心理机制	088
第七章	惩罚的概念、类型、理论与作用	093
第一节	惩罚的概念	095
第二节	惩罚的类型	095
第三节	惩罚的理论	099
第四节	惩罚的作用	101
第八章	惩罚影响合作的动机——信任	107
第一节	惩罚作用的争论	109
第二节	惩罚强弱的作用探讨	114
第三节	惩罚通过信任影响合作	115
第九章	心理学中的信任与合作	123
第一节	家庭环境、亲子依恋与人际信任	125
第二节	进化心理学与信任和合作	130
第三节	教育心理学视野下的合作	135

第十章 信任与合作的本土化	143
第一节 儒家思想与信任的关系	145
第二节 社会转型与熟人信任的局限	147
第三节 从熟人信任到制度信任	150
参考文献	156
附录	174
附录一 人际信任与信任行为的实证研究	174
附录二 社会价值取向、感戴与信任的实证研究	183
附录三 惩罚强弱影响合作的实验研究	188
附录四 惩罚、信任与合作关系的实验研究	194
后记	203

第一章

社会困境的内涵和类型

第一节 社会困境的内涵

社会困境也称为“社会两难”，描述的是一种个人利益和群体利益存在冲突的情景。当群体中所有成员都按照自己的利益行动，而忽略群体的利益时，最终会导致个人利益的最大损失。社会困境的产生要素包括以下三个：①由个人组成的群体；②在他们之间存在的共同利益；③在共同利益和个人利益之间存在的潜在性冲突。社会困境中的参与人之间并非一种“你死我活、非此即彼”的冲突状态（零和博弈），而是存在着一种共同利益（非零和博弈）。参与人之间不是势不两立的敌人，而是存在疑虑和分歧的伙伴，试图通过各种策略的博弈以寻求双赢的结果。

社会困境中的参与人有两种决策行为，一种是合作，另一种是不合作。合作是指参与人为了完成互相依赖的博弈任务，一方愿意顾及对手的感受和收益，达到互惠共赢的结果。与合作相反的概念是不合作，也称为背叛、背信、“搭便车”行为，即参与人在决策时只考虑个人利益的最大化，不顾及他人的利益和感受。

Dawes（1980）认为社会两难具有以下三个特征：①无论其他人做出什么选择，参与人采取不合作的选择总是比合作的选择更有利可图；②与合作的选择相比，不合作的选择对他人总是有损失的；③不合作的选择对他人造成的损失总计大于参与人个人的收益。

合作是指个人与个人、群体与群体之间为达到共同目的相互配合的一种联合行动。社会两难中的合作特指在一般合作定义的基础上，参与人为了完成互相依赖的博弈任务，一方愿意顾及对手的感受和收益，达到互惠共赢的结果。对参与人而言，只有具备共同的目标（至少是短期目标）、统一的认识和共同遵守的社会规范、互相信赖的良好气氛，以及合作赖以生存和发展的物质基础，合作才有可能得以实现。

社会两难问题的研究最早来自博弈理论，自20世纪60年代起，作为利益冲突条件下最优策略问题的理论，传统的博弈理论被介绍到社会和行为科学中，它认为人是理性的和利己的，在决策时每个人都会追求约束条件下收益的

最大化；同时，人的行为又是互相影响的，每个人的得益不仅取决于自己的策略选择，还取决于博弈中其他的参与人，即“知己”与“知彼”是策略选取的起点。

简单地讲，博弈论的基本概念包括参与人、行动、策略、信息、支付函数、结果和均衡。参与人是在博弈中选择行动以最大化自我效用的决策主体；行动是参与人的决策变量；策略是参与人选择行动的规则，它告诉参与人在什么时候选择什么行动；信息是参与人在博弈中的知识，特别是关于其他参与人策略或行动的知识；支付函数是参与人从博弈中获得的效用水平，它是所有参与人战略或行动的函数，是每个参与人真正关心的东西；结果是指博弈分析者感兴趣的要素的集合；均衡是所有参与人的最优战略或行动的组合。

最有代表性的社会两难情景是“囚徒困境”，另外，经常讨论的还有以下一些形式的社会两难情景：资源两难、公共物品两难、独裁者博弈、信任博弈、最后通牒博弈、礼品交换博弈、第三者制裁博弈等。

在经济学中，“理性人”是根据在经济社会中从事经济活动的所有人的基本特征的一般性抽象个体，其基本特征是：每一个从事经济活动的人都是利己的。也可以说，每一个从事经济活动的人所采取的经济行为都是力图以自己的最小经济代价去获得自己的最大经济利益。“理性人”也被称为“经济人”。

“理性人”假说认为人在本质上是自私自利的，在信息充分的前提条件下有如下特征。

- 1) “理性人”具有关于他所处环境的完备知识，而且这些知识即使不是绝对完备的，至少也是相当丰富、相当透彻的。
- 2) “理性人”有稳定的和条理清楚的偏好。
- 3) “理性人”有很强的计算能力，能估算出每种选择的后果。
- 4) “理性人”能使其选中的方案自然达到其偏好尺度上的最高点，即“理性人”总能选择最优。

传统的博弈理论有两个目标：①通过描述和发展博弈实验，从理论上寻找利益的分配方式；②通过博弈分析，发现解决利益合理分配的方法。

博弈论有两类：①合作博弈——研究人们达成合作时如何分配合作得到的收益，即收益分配问题。合作博弈强调的是集体主义、团体理性（collective rationality），是效率、公平、公正。②非合作博弈——研究人们在利益相互影响的局势中如何决策使自己的收益最大，即策略选择问题。非合作博弈是指人

们的行为在相互作用时，当事人不能达成一个具有约束力的协议，强调的是个人理性、个人最优决策，其结果可能是有效率的，也可能是没效率的。

早期的纳什均衡（Nash equilibrium）理论揭示了博弈均衡与经济均衡之间的内在关系，是对冯·诺依曼和摩根斯特恩的合作博弈理论的重大发展，奠定了现代非合作博弈理论的基石，后来的博弈研究基本上是沿着这条主线展开的。

所谓纳什均衡是指假设有 n 个人参与博弈，给定其他人策略的条件下，每个参与人选择自己的占优策略，从而使个人效用最大化。所有人的个人占优策略构成的一个策略组合就是纳什均衡，也叫“非合作均衡”。因为每一方在选择策略时都没有办法沟通信息，他们只能选择对自己最有利的策略，而不考虑社会福利或任何对手的利益。也就是说，这种策略组合由所有参与人的占优策略组合构成，没有人会主动改变自己的策略以便使自己获得更大利益。所谓帕累托最优（Pareto optimality）是资源分配的一种状态，指在不使任何人情况变坏的情况下，使得至少一个人变得更好，即帕累托最优状态就是不可能再有更多的帕累托改进的余地。它是公平与效率的理想状态。

纳什均衡对“看不见的手”的原理提出挑战。纳什均衡发现该原理的一个悖论，即两个囚徒从个人利己目的出发，其结果是损人不利己，既不利己也不利他。纳什均衡提出的悖论实际上动摇了西方经济学的基石，从纳什均衡中可以悟出一条真理：合作是有利的利己策略，但它必须符合以下黄金律——按照你愿意别人对你的方式来对别人，但只有他们也按同样方式行事才行，也就是中国人说的“己所不欲，勿施于人”，但前提是“人所不欲，勿施于我”。

从经济学的“理性人”假设来看，社会困境是个体理性导致集体非理性的情境。对个人来说，也许个人利益最大化更有吸引力，但是如果所有的人都选择个人利益最大化，其结果是每个人的收益都将低于群体利益最大化时的收益。也就是说，社会两难是这样一种情景：不管其他人怎么做，每个人都按照自己的利益行事，然而每个利己的决策对涉及的其他人都会产生一个消极的结果。

市场经济学谈到资源在市场中的配置时，常常提到市场中好像有一双“看不见的手”，每个人都试图应用自己的资本，来使其产品得到最大的价值。一般来说，他并不企图增进公共福利，也不清楚增进的公共福利有多少，他所追求的仅仅是个人的安乐、个人的利益，但当他这样做的时候，就会有一双“看不见的手”引导他去达到另一个目标，而这个目标绝不是他所追求的东西。他