

TRUST:

MAKING AND BREAKING
COOPERATIVE
RELATIONS

信任:

合作关系的建立与破坏

郑也夫◎编

信任理论文丛

中国城市出版社

TRUST:

MANAGING TRUST

CONTRACT

RELATIONSHIP

信任：

合作关系的建立与破坏

陈怡人 著



管理理论丛书

中国社会科学出版社

信任：合作 关系的建立与破坏

郑也夫 编
杨玉明 皮子林 等译

中国城市出版社

图书在版编目 (CIP) 数据

信任: 合作关系的建立与破坏/郑也夫编. —北京:
中国城市出版社, 2003.1

ISBN 7-5074-1479-5

I. 信… II. 郑… III. 社会问题—文集

IV. C913-53

中国版本图书馆 CIP 数据核字 (2003) 第 001120 号

责任编辑	何玉兴		
封面设计	耀午书装		
责任技术编辑	张建军		
出版发行	中国城市出版社		
地址	北京市朝阳区和平里西街 21 号	邮编	100013
电话	84275833	传真	84278264
电子信箱	citypress@sina.com		
读者服务部	84277987		
经销	新华书店		
印刷	北京集惠印刷有限公司		
字数	217 千字	印张	9.5
开本	850 × 1168 (毫米)		1/32
版次	2003 年 2 月第 1 版		
印次	2003 年 2 月第 1 次印刷		
印数	0001—5000 册	定价	18.00 元

本书封底贴有防伪标识。版权所有, 盗印必究。

举报电话: (010) 84276257 84276253

目 录

形式结构与社会现实	伯纳德·威廉姆斯/ 1
合作与信任的生物进化	帕特里克·贝特森/15
个体、人际关系与信任	大卫·古德/34
作为商品的信任	帕萨·达斯古普塔/57
信任与政治行为	约翰·邓恩/89
熟悉、信赖、信任：问题与 替代选择	尼克拉斯·卢曼/116
信任的三大讽刺	杰弗里·霍桑/135
信任毁灭及其经济后果：以 18 世纪的 那不勒斯为例	安东尼·帕格顿/157
信任、凝聚力和社会秩序	欧内斯特·戈尔尼/176
黑手党：不信任的代价	迪戈·甘姆贝塔/196
亲戚、合同和信任：一非洲城市贫民窟的 移民经济组织	基思·哈特/218
既非朋友亦非陌生人：法国工业转包的 非正式网络	爱德华·H·罗伦兹/243
我们能信任信任吗	迪戈·甘姆贝塔/264
译后记	郑也夫/298

形式结构与社会现实

伯纳德·威廉姆斯 (Bernard Williams)

王琳予 译 郑也夫 校

在本文第一部分，我对博弈理论及其社会现实中的运用提出疑问，并作分析。我强调，不从单一回合的博弈也不从重复博弈角度，而从社会存在的普遍动机和倾向角度去思考才有价值。第二部分我思考包含在“合作动机”中的某些复杂因素，以及“合作动机”观念与信任观念两者的联系。在第三部分，我对社会合作框架得以维持的必要而充分的动机，提出一些非常具有普遍意义的假设。

一

我从经典囚徒困境的观点谈起。 X_1 代表囚徒 X 招供， X_0 代表 X 不招，标准博弈（囚徒困境）中两位参与者 A 与 B 的优先选择排序如下：

A: $a_1b_0 > a_0b_0 > a_1b_1 > a_0b_1$ （“标准博弈”）

B: $a_0b_1 > a_0b_0 > a_1b_1 > a_1b_0$

显然这种排序导致众所周知的结果：因为对每个参与者，招供绝对胜过不招，所以他们都招供并都得到比不招更糟糕的

结果。

也可能有其他的优先选择，森（Sen：1974）曾经探讨过其中部分。要加思考的是那些被森加作“有保证的博弈”情形：

$$A: a_0b_0 > a_1b_0 > a_1b_1 > a_0b_1 \quad (\text{“有保证的博弈”})$$

$$B: a_0b_0 > a_0b_1 > a_1b_1 > a_1b_0$$

森（Sen：1974：78）写道：双方都不招的约定在“有保证的博弈”中不需要任何强化，在“标准博弈”中则成为关键所在。这样讲可能有些误导。在“有保证的博弈”中双方排在第一位的偏好选择是“你不招我也不招”，这样讲没错；但是如森所说，双方都需要对方的保证，约定履行得怎样还要看如何获取对方的保证。这道出在对制裁可能所起作用的认识上存在着一个重要分歧，一个与政治学理论有关的分歧。制裁加码可能改变最初利己的参与者的效用，也可能为最初愿意合作的参与者提供保证。后一种可能性意义重大，和我将要深入探讨的一个观点相关，那就是：合作倾向典型地受成本影响。

保证也能够通过其他办法去获得。所需保证的复杂性是众所周知的。双方都必须清楚对方是在“有保证的博弈”中选择对策，必须知道对方对他（第一人）的选择了然于心；若办不到这一点，双方都认为有被对方出卖的风险，就会揣度如何规避，如何抢先制胜，因此就可能违背初衷而放弃双方约定。这是因为任何一方都有如下的偏好选择：

我招你也招 >> 我不招你招

缺少这点认识，或缺少对参与其中的某些情况的了解，合作就会解体。博弈的基本理论认为，这样的认识是一个先决的假设，因为每个参与者都被假设为能够获知完整的相关信息，也都具有完全合理地利用信息的能力。

不过，近期的研究更乐意去削弱这些假设。现实中存在着几种局限，例如：

现实局限一：

- (1) 其他人的选择偏好及对或然性的估计，人们不能够尽如人意地获取到这两方面信息；
- (2) 上述局性 (1) 得不到充分理解；
- (3) 受各种因素影响，获取这些信息的可能性小而且代价高昂。特别困难的是：任何现实的探询步骤本身就可能引起参与者偏好改变，破坏了信息，引发更多疑问，使得问题更加扑朔迷离；
- (4) 除了认知的缘故，社会因素也给推测带来不容小视的局限。

现实生活中参与者在认知上遇到的局限，我把这类生动事实统称为“现实局限一”。

关于现实局限一之 (4) 以及纯粹的认识局限如何被超越，法国战后的共谋史实（威廉姆斯：1970）算得上生动详尽的一例。这类不确定性具有的巴洛克风格的复杂结构体现在，一方行动旨在产生这样的印象，即另一方行动之目的是生成第一方加入同谋的印象，这一结构必然导致了没有人真正了解人们在做什么；而且，缺乏了解还将影响到已掌握的情况。因为如此情形是由当事人的意图勾勒出来的，而对他人意图的把握愈来愈不明朗使当事人自己的意图受到局限，于是发展到谁也说不清“人们正在干什么”。

森还提到如下第三种优先选择排序，他称之为“重视他方”的对局：

$$A: a_0b_0 > a_0b_1 > a_1b_0 > a_1b_1$$

$$B: a_0b_0 > a_1b_0 > a_0b_1 > a_1b_1$$

森认为，有了这种优先选择方案，任何参与者不再需要得到保证，因为无论 A 或 B，不招绝对是最有利的策略。不过我们也由此看到，传统理论中还夹杂着另一类非现实的人为假设：传统理论结构既包含一种优先选择的一次性表达，又包含不受学习影响的一种偏好的重复表达。这就是现实局限二：

与合作相关的优先选择在发生改变，尤其是受到有关“对方给予的各项保证的可靠性”这类信息影响的情况下。

于是人们可能从“重视他方”的“你我都不招”的对策开局，却因频繁遇到“我不招你却招”的情形而受挫。

在经典的非重复性（或只是微小重复）博弈的理论中，当交易被看作是选择偏好的表达时，我们可以单纯从后果带来的效用来理解这些偏好。这些把博弈按其特点划分为“标准博弈”、“有保证的博弈”和“重视他方的博弈”的优先选择，可能被描述成“利己的”、“合作性的”、“利他的”等等，但这样划分只能提供出一个心理画面而已，合作要素是受一组结局限定的选择偏好。然而现实中，选择偏好毕竟反映了参与者现实的心理状态，尤其反映出他的倾向；一个既定场合中是否同他人合作的既定选择就是几种倾向和态度的函数，——是冒险？是合作？是同这类企业打交道？同这方还是同他所属的团体打交道？等等。对参与者倾向的这些分类事实将对其他参与者的信念起重要作用，并影响参与者本人作出的决定。

艾克斯罗德（Axelrod: 1984）进行的不同策略的反复比试饶有趣味和意义，但只与社会中人类的合作问题间接相关。所有参赛者都在“囚徒困境”中对局，所以他们对优先选择的排序是一致的。不同的只是策略，这些策略致力于由他们的优先选择决定

的尽可能多拿分的目的。不同的参与者是不同的策略，分数落到不同的策略上。于是，该系统模拟的是某一条件行为模式如何生存与取胜；对它最自然的解释是利用进化论学说蓄意而为的。^①一个参与者在社会互动中，他的策略和选择偏好比如对合作的渴望，都可能受到条件影响，也可能被经验修正。

当然，一个人在既定场合参与合作，不一定是因为他怀有利他的合作动机。我们已经注意到，而且文学作品也老在提醒我们，来自制裁的威胁施用到利己参与者身上可能产生同样的结果。除了这种可能性，毕竟还存在着不是利己的合作动机，但这种动机要比“重视他方”的优先策略的最简单的心理概括所代表的那类事物受到更多限定。受条件影响的种种普遍的合作倾向存在，且尤其对成本敏感：如果代价无多，参与者会选择合作，而超过这一阈值他会放弃合作。这一阈值本身会随着时间推移发生改变。不仅如此，在不能对其他参与者的倾向了如指掌的时候，人们也可能冒险参与合作，前提是万一受骗，蚀本要低于一定的阈值。由此我们发现了现实局限三：

合作倾向受到成本影响。

讲求成本不一定就是利己表现，这一点非常重要。事实上，有受挫经历的参与者可能出于利己的理由而拒绝“重视他方”的那种最简单的心理法则（即他总是抱着以“重视他方”策略来谋取结果的心理倾向）。拒绝的理由不光只用“利己行为”来解释，比如（特别是发现成本之类无法接受的时候）还可以把拒绝当作是对自我保护的权利的要求，或是因为对其他团体、家庭、友人承担着义务，等等。

① 参见本书P·贝特森(P. Bateson)。从艾克斯罗德进行的竞赛得到的结果中能合理地发现一个让人不疑的特征，那就是所有成功的对策无一不是“善良的”，它们从不背叛。

二

前面的讨论经常提到合作的动机(或曰倾向),在进入第三部分思考合作动机的不同类型之前,我们应进一步考查什么是合作以及合作的动机,由此将提出若干假设作为下一步讨论的基石。我认为:

两个参与者合作是在他们对共同参与的行动带来的后果共担风险的时候,并且其中至少一方的一个行动发生在不受对方直接控制之下。

(这样的描述完全可以推广到涵盖两方以上的合作:言明一方要合作或参与合作,同时存在着与之合作的对象。)

此定义下,两方合作的情况必然涉及到至少有一方依靠另一方,我也愿意这样讲,有一方是依赖方。当然还有些风险联合(风险共担)自然被叫做“合作”,却没有被归于此定义之下,因为风险联合中每个人都处于每个人的直接控制下,任何一方不尽职责,都会马上被发现。不过提到这些对我们目前的讨论没有意义,现在讨论的全都是一方依靠另一方的情况。(这方面的一个重要例子是,一方先于另一方采取关键的举动,但如传统“囚徒困境”所述,这其实是不必要的。)虽然不太直接,但就是这些例子最终把我们引到了信任的概念。

合作是一种对称关系:如果 X 同 Y 合作, Y 也同 X 合作(若这样看显得牵强,请把“合作”理解为“参与到与…的合作性风险联合中”)。“依靠”则是一种非对称关系:如果 X 依靠 Y——意味着 X 能否从风险联合中得到他所想^① 取决于 Y

^① “得到他想要的”而不是以他人“回报”作参考,要强调参与者加入合作性风险联合的目的不一定是利己的。一个重要的案例是参与者从风险联合中得到的满意一定暗含着风险联合的成功——例如,因为参与者想要的是通向成功的风险的联合,而不是一旦风险联合失败他能从中得到什么。如果所有参与者都希望有这类满意,那么实际上他们就是相互依靠的。一个社会的存在涉及很多这类情况。

在其中的表现——Y可能依赖X，也可能不依赖X。

有了这些定义，我们就可以把不同类型的动机清晰地区分开，这些动机可被称做“合作的动机”。最宽泛地讲，“合作的动机”包括任何一种进入到风险联合，进入到将与人合作的情形的动机。然而，在不同场合、不同条件下，不是所有促使人们走向合作的动机都能取得同样结果。具体说，在某些合作中，动机允许参与者是依赖方；而另一些合作中，即使有目的的合作动机存在，也将只在参与者不充当依赖方的前提下促使其进入合作。囚徒困境及近似问题的一个来源在于事实上纷繁复杂的动机并非如出一辙。

社会理论要作出的回答：需要存在什么样的动机才会有合作？这也是第三部分着力探讨的问题。很快就会明白这一点，如果进到合作性风险联合的普遍动机仅仅是谁也不能（心知肚明地）依靠谁，就不会出现任何合作。

这个不证自明且与社会相关的命题，有必要同另一个放在具体条件下才成立且只对个人的主张区分开来。后一主张认为，如果不能处于依赖方位置，或者处于非依赖方位置且时时背叛的话，个体参与者就不会希望进入合作性的风险联合。在参与者及其倾向都众所周知的团体中这样讲无疑是对的，放在其他团体中就不一定了。我的问题是：普遍而言，什么类型的结构可能支持合作？一个可能的回答是：产生合作的惟一途径是把合作局限在个人倾向和性格都彼此熟悉的人当中（彼此间的信赖可以称做“深厚的信任”），这个答案多少让人对现代生活感到惆怅；不过现在这只是一个答案，还有其他可能的答案回答“人们要去合作，究竟需要什么样的动机存在”。

要想持续地去实践合作，人们就得受这样那样的动机驱使进到依赖方的位置，这是一个浅显的命题，很容易往下推进两

步。首先，除非有几分确信对方作为非依赖方不会背叛，人们通常自己不会背叛。此观点的一个非常抽象的概括就是前面已提到过的博弈理论结构的核心。信任的概念也在于此。合作需要信任，依赖方需要某种程度上确保非依赖方不会背叛，我已讲过，这种保证不一定是人与人之间“深厚的信任”，但一定是具有某种形式的信赖。接下去阐发出第二步主张，人们一般不会信任对方到足以产生合作，除非他们的保证在某种程度上有很强基础，也就是说，除非处于非依赖方位置上的人们受到这样或那样的一般性动机驱使而不去背叛。

这两步推论牵涉到经验的主张，确实在有些案例中一个甚至两个推论都会被推翻，但是毕竟大部分内容经过了重重验证，将其作为探讨信任与合作问题的一个基础是合理的。我将接受这些主张，由此开始对合作动机的探讨。我将要谈到合作的种种动机及动机可能的由来，有时候是不得不成为依赖方的动机，有时候是非依赖方不会叛变的动机。这是两类截然不同的事物。但是，一旦我们抛开单一回合博弈不谈，彻底地思考社会合作实践的意义，如果接受刚刚提到的经验假设，我们就完全有理由把这两类事物合为一类来研究，因为在极大程度上，两类事物通常情况下是共存的——这里讲“共存”指“存在于同一社会”，但不一定“发生在同一人身上”。

当思考各种动机以及相应地各种制度设置如何发挥作用或解体时，把那些促使个人参与合作或拒绝合作的形形色色的倾向纳入考虑，相当重要。本文第二部分我致力于探讨对该问题的若干假设。这些假设仍然相当地粗略，意在由此能引出系统的思考及探询，而不是就此给问题画上句号。

三

我从一句格言展开论述，人们把它叫做休谟的格言：

动力只源于动机。

有人会认为这是个同义反复，但我正是这样理解的。这句话和我们正进行的讨论相关，意思是如果行动者的目的是为了促成某种共同期待的结果，那么该行动只能够用行动者拥有的某种倾向和愿望（例如为什么只那样做）来解释。行动者的行动不能够只拿行动促成那样的结果这一事实或单用行动者对结果的知晓来解释。由此推出：希望人们去追求某种结果，就务必使带链该结果的适当动机现实地存在。什么类型的动机能使合作的可能性稳定下来？

首先，介绍一下宏观动机与微观动机的差异将有所帮助。对此差异的概述还需要大量推敲和斟酌：我希望下面的论述对目前的讨论来讲是充分的。假设参与者经常有合作的举动，并且这类举动的合作性带有一种蓄意而为的特征。我们可以从不同层次来解释这些举动。对于某些参与者，解释的方式可以用合作的普遍动机；把这些举动当作合作性加以描述成为对合作者在所有这些场合中表现出的某一动机加以刻画的一部分。该行动者拥有趋于合作的宏观动机。

合作性的宏观动机不一定就是利他的或道德的。利己的宏观的合作动机可能存在。霍布斯式结构中有这种情形。这样的合作动机是利己的，因为本质上出于对君主权力的畏惧。不过，这样的合作动机又是宏观的动机，因为一般而言君主权力蓄意要带来的就是合作动机。从利己的角度考虑，参与者在许多这样的场合选择了合作不是偶然；准确地说，他清楚自己带着利己的理由去做的事情就是合作。

非利己的、宏观动机可以在某种伦理或道德倾向中被找到，如承认忠诚是一项普遍义务，承认对要求合作的国家政权的义务。它可能以参与者与宗教规范的关系为基础，尽管有些严厉的宗教道德完全用奖赏和惩戒来表达，可能让我们又回到利己的选择——具有神圣君权的霍布斯系统。

微观的合作动机是指在既定场合（一个或数个）没有包含任何合作的普遍动机时产生的去合作的动机。非利己的微观动机可能存在与既定个人的友好关系当中，对其自然的表达包括与她合作的愿望。进一步说，我可能有进入这种友好关系的倾向，所以任何时候我都愿意合作；但即使这样，我可能仍然没有走向合作的宏观动机，该宏观动机与倾向于进入友情并从中趋向合作相反。

最后，利己的微观动机是颇为自明的：在这种情形中，当我把它看做是对我的私利有益时，我有合作的动机，但这并不是如同在霍布斯系统中一般地受到动机驱使去合作。这并不排除如下可能性：我可能定期发现自己处于这样的情形当中，或是存在某种结构以至于我定期发现自己处于这样的情形当中，产生出这种结构可能还是“无形的手”取胜了。然而，这类动机将有一个以某种纯粹形式体现出的标志：从这次合作到下次合作不再有动机的推动。我很乐意下一次与 X 合作，因为我欣然相信下一次与 X 的合作对我有利，但是并不存在合作的普遍倾向，甚至连与 X 合作的倾向也没有，就是这样，如果既定场合的特殊情况不能使合作对我有利，我就不与 X 合作。

如果现在从动机的这些类型着眼提出问题：社会需要什么以维持合作的结构？若干假设结论看似可信。

一、宏观动机不能独自发挥作用

(一) 利己的宏观动机不能独自发挥作用：它要么收效甚微，要么招人厌嫌，甚至二者皆有

因为在此选择下，合作的普遍动机与个人对自身利益的感知直接联系在一起，对发现破绽的预期这类情况极端敏感，又因为这一事实本身是常识，所以先发制人的行动（如在经典的囚徒困境中）不断受到掣掇。故而这系统往往会解体，趋于不稳定。如果人们的宗教意识能复苏的话，那么宗教的说法（尽管少有不被憎恶的时候）拿来解释也许更能奏效，然而在声称相信它的社会中的被观察到的行为并不能证明它。可能是这样，百分之百发现比率的确定性已被法庭实践的不确定因素抵消了。

(二) 非利己的宏观动机不能独自发挥作用

这类倾向最纯粹的表现形式是前文讲到的“重视他方”的优先选择的最一般的心理概括，尤其是因为受到现实局限二和三的影响，这类倾向不稳定。前面已讲到，这类倾向可能受条件影响而变得更加复杂：有人可能认为某些条件满足了，就可以去合作；反之就不该合作。但这些防范需要暗示了参与者还有其他的动机，尤其是利己的动机。如果合作不在某种程度上维护当事人利益，那么合作的实践就会在其他动机面前松动。非利己的宏观动机不能独自发挥作用的另一缘故，不是合作中潜在的敌人，而是合作提供出的各种奖赏。很多时候合作令参与者满意是因为合作暗示了合作但又不集中在合作上，除非有大量类似的合作存在，合作不会对人们具有非常普遍的吸引力，合作还可能在怨怒中瓦解。这构成对单纯以利己主义为基础的社会的设置的一种特殊的危害，这些利己主义受到少许道德准则的修正，比如对信守诺言的要求。

二、利己的微观动机不能独自发挥作用

缺少某种一般性结构，利己的微观动机中只能让我们再次陷入囚徒困境。因此，如果利己的微观的动机能说明一切，我们就需要一只“无形的手”的结构，使背叛的趋势被某种利己的微观的动机所抵消（同时还使得它充分地被知晓，等等）。但是我们需要靠异乎寻常强大而又难以置信的功能主义的假设，才能相信“无形的手”能够自动调节与自我维系，与此同时，大量社会实验的无情事实带给我们以丰富证据却显示，自动调节、自我维系并不存在。

结论一第1条和结论二合在一起，包含着：

三、利己动机不能独自发挥作用

由以上三条结论得出，如果有独自发挥作用的动机存在，那一定是非利己的微观的动机。这听上去好似“深厚的信任”——群体内成员依据共知的倾向、个人关系等等而相互信赖。要了解信任如何扩展到很小的社会之外并非易事。它惟一可见的扩展路径是通过识别某一特定人群、某一政府或某一精英集团，在这些群体当中已投资了“深厚的信任”，这些群体的活动或规则致力于在整个社会中维持合作性活动；但是，如果这完全只是想像，我就无法看清楚，离开了神秘化的人格魅力和传统主义，“无形的手”怎么可能成为我们制度安排的基础？事实上显而易见的是：

四、对我们而言，非利己的微观动机不能独自发挥作用

“我们”栖息在现代社会，陷入千丝万缕的社会关系当