

# Globethics Repository

The logo for Globethics, featuring the word "Globethics" in white sans-serif font on a blue rectangular background.

## “囚徒悖论” ["Prisoner's Dilemma"]

This page was generated automatically upon download from the Globethics Repository. More information on Globethics see <https://www.globethics.net>. Data and content policy of Globethics Repository see <https://repository.globethics.net/pages/policy>.

Item Type	Article
Authors	陈, 真
Publisher	云南大学
Rights	With permission of the license/copyright holder
Download date	2026-07-05 01:09:36
Link to Item	<a href="http://hdl.handle.net/20.500.12424/185670">http://hdl.handle.net/20.500.12424/185670</a>

# 陈真：“囚徒悖论”：道德的合理性和国民的道德教育

陈真

“囚徒悖论”：道德的合理性和国民的道德教育

陈真

南京师范大学哲学系

关键词：道德行为；合理性；囚徒悖论；道德要求；道德教育

摘要：通过对决策理论中“囚徒悖论”问题的考查，找出道德合理性的必要条件和充分条件，并在此基础上进一步证明，除了法制和制度性建设外，我们同时也需要进行内容合理的、有效的道德教育，以提高全体国民的道德素养。反过来，全体国民的道德素养也会为道德观念的合理性创造条件、奠定基础。

中国传统的哲学家和绝大多数的西方哲学家都认为道德包含着利他主义的因素或者对他人利益的考虑。[1] 但一谈到利他主义，人们就想起“毫不利己，专门利人”，好像一个有道德的人就一定要做到“毫不利己，专门利人”。这种道德要求对正常的和平环境中的普通人来说，其不合理性是显而易见的。其结果造成道德教育的低效，甚至无效。另一方面，由于这种道德要求的不合理性造成一部分人认为合理的道德要求完全没有利他主义的成分，或者认为“利他”（如代表整体利益的行为）和“利己”是一回事。[2] 我认为将道德的要求等同于“毫不利己，专门利人”，或者将代表整体利益的道德行为等同于利己的行为都是不对的。道德要求需要讲求合理性。要找到这样一种合理的道德，我们需要找出其合理性的基本条件。本文打算通过对决策理论中“囚徒悖论”问题的考查，找出道德合理性的必要条件和充分条件，并在此基础上进一步证明，除了法制和制度性建设外，我们同时也需要进行内容合理的、有效的道德教育，以提高全体国民的道德素养，反过来，全体国民的道德素养也会为道德观念的合理性创造条件、奠定基础。

## 一、关于道德合理性的概念

中西方哲学家一般都认为凡是道德的也都应该是合理的。合理性包含着认识的合理性和实践的（行动的）合理性两个方面。人们的认识是否合理主要取决于我们究竟有无充分的理由相信某个命题为真。评判认识合理性的最终根据是客观的事实或有关客观事实的证据。而实践的合理性则主要取决于我们究竟有无充分的理由采取某种行动或做某件事。评判行动理由的最终根据是行动的目的一定要有价值。没有价值，便没有真正的行动理由。价值大致可以分为对自己（行动者）的价值和对他人或社会整体的价值。[3]

本文所要讨论的道德合理性主要指的是实践的合理性，即道德行为（包括道德规则及其实践）的合理性。道德的行为是否合理取决于我们究竟有无充分的理由做有道德的事情。如果有，道德的行为就是合理的。如果无，道德的行为就不合理。人们通常认为我们当然有理由做有道德的事情，但这里有两方面的问题需要回答：第一，如果我们有理由做有道德的事情，我们有什么样的理由做有道德的事情？人们可能认为我们有理由做有道德的事情是因为道德的行

为代表了我们的整体的、长远的利益。那么，道德的行为和要求究竟怎样代表了我们的整体的、长远的利益？回答这方面的问题有助于我们找出合理道德行为的必要条件。第二，即使道德的要求代表了我们的整体的利益，但并不等于代表了我们的个人利益。当代表整体利益的道德要求和我们个人利益发生冲突时，我们还有充分的理由按照道德的要求办事吗？回答这方面的问题有助于我们找出合理道德行为的充分条件。

上述这些问题涉及到整体利益和个人利益的关系问题。由于“囚徒悖论”将整个社会简化成二个人（有时是多个人）之间的关系，在这个最简单、最基本的“社会”中，整体利益（两个人的共同利益）和每个人的个人利益之间的关系能够表现得最清楚、最明白。故考察“囚徒悖论”的问题可以有助于我们回答上述问题。

让我们先来看看两种不同的“囚徒悖论”情景。

## 二、一次性“囚徒悖论”和重复性“囚徒悖论”

众所周知，一次性“囚徒悖论”是博弈论中的一个特殊情景。囚徒或行动者必须在仅有的两种行为中选择最合理、最有理由的行动。“囚徒悖论”所显示的问题是：每位囚徒从本身的个人利益出发，都最有理由坦白交代，出卖对方。然而，双方都出卖对方的结果比双方都合作即守口如瓶的结果更糟，即对双方都不利。为了避免这种的结果，双方合作似乎就成了唯一的出路。然而，合作即守口如瓶的问题是：如果另一方不合作，则合作总是对合作的一方不利，因而这种合作就失去了合理的基础。如何才能确保对方合作，从而使己方的合作成为合理的行为，以避免对双方都不利的结果？这一问题称之为“确信问题”。

这一问题在重复性的“囚徒悖论”情景中似乎可以避免。在重复性的“囚徒悖论”中，决策的情景只是一系列的“囚徒悖论”情景中的一环，当事人或行动者并不知道究竟有多少类似的情景还会发生。举例来说，假定有两位农场主丙和丁，处于自然状态下（即处于没有警察、军队等维持治安的国家机器的状态下）。由于他们经常受到贪婪匪徒的威胁和袭击，丙和丁达成一个防御协定：当一方遭到袭击时，另一方应该及时援助。假定丙方遭到匪徒袭击，丁方面临两种选择：或遵守协定，援助丙方，或违反协定，按兵不动。作为一个有理性的人，丁应该怎样做才最合乎理性呢？假定他们从行动中所得到的益处可以用期望功利值来衡量，则其行为的后果以及可能给各方带来的利益可以用下图来表示：

		丙	
		遵守协定	违反协定
丁	遵守协定	6	10
	违反协定	0	1

图中数值代表了定约人不同行为（遵守或违反协定）对各自产生的期望值（即各自可能得到的好处）。数值越高，则益处越大，因而定约人更有可能采取相应的行动。其中每个方框中左下角的数值代表了丁可能得到的益处，每个方框中右上角的数值则代表了丙可能得到的好处。

如图表所示，如果丙丁两方都遵守协定，则各自所获的期望值为 6。如果丙方违反协定而丁方遵守，则丙方所获的期望值为10，而丁方则为 0，反之亦然。如果双方都违反协定，则双方所得到的功利值为 1，低于双方都遵守协定所得到的功利值 6。[4] 一般说来，我们将遵

守协定，援助对方看作是道德所要求的，而将违反协定看作是不道德的。那么，在“囚徒悖论”或类似的情景中，我们究竟有无理由做道德所要求的事情呢？

### 三、道德合理性的必要条件

在农场主的例子中，如果我们将其看作是一次性的“囚徒悖论”情景，行动者似乎并无理由做有道德的事情。因为，不论另一方采取什么行动，违反协定的好处总是大于遵守协定的好处。如果对方遵守协定，则违反协定可使己方获得最大的好处，即获得功利值10。如果对方也违反协定，则己方违反协定可以避免最糟糕的情形，即一无所获的情形出现。尤其是，如果己方后于对方行动，则违反协定肯定最符合己方的利益。

但有人认为，在一次性“囚徒悖论”情景中，如果假定对方也能做同样的事情，行动者也有理由做有道德的事情。理由就在于双方都做有道德的事情比双方都不做有道德的事情对双方都有利得多，也就是说道德的行为和要求可以避免双方仅按照个人利益行事所带来的不利结果，因而代表了双方的整体利益。这里有几个问题需要注意。第一，这里的“整体利益”并不等于行动者的个人利益，因为违反协定对个人更有利。如果从整体利益出发，行动者必须放弃对个人利益无限制的追求，必须“牺牲”某些个人的利益。第二，正是在这一意义上，从整体利益出发遵守协议的行为包含有“利他”的因素。第三，“整体利益”确实构成了行动者做道德事情的必要的理由，没有这一理由，道德的要求和行为就失去了合理性的必要条件。但第四，“整体利益”并不能构成道德行为的充分理由。因为在一次性“囚徒悖论”的情景中，行动者无法知道对方想干什么。在不知道对方是否是一个利己主义者的情况下，行动者并无充分理由相信做有道德的事情将会有利于双方。也就是说，无条件的利他主义行为未必就一定代表双方的整体利益。

如果我们将一次性“囚徒悖论”的情景改变成重复性的“囚徒悖论”情景，则情况似乎会有所不同。在重复性的“囚徒悖论”情景中，行动者有两条理性的原则可供选择。第一条原则主张行动者应采取马上可以带来好处的行为，即所谓“短期化的行为”。第二条原则主张行动者应采取能带来长远利益或好处的行为。主张第一条原则的人可称之为“急功近利的理性主义者”。主张第二条原则的人可称之为“有远见的理性主义者”。

按照第一条原则，上述例子中的丙丁就会违反协定，因为违反协定能带给他们“立竿见影”的好处。当对方违反协定时，己方违反协定可以得到期望功利值1，而不是0。当对方遵守协定时，则违反协定可给己方带来10个期望功利值的好处，而不是6个。问题在于，如果双方都是“急功近利的理性主义者”，则双方都不可能避免所谓“次佳化”的结果，即每个行动者只能得到1个功利值。相比之下，如果所有的行动者都遵守协定，则每个行动者都可以得到6个功利值。

按照第二条原则行事，丙丁则愿意合作，遵守协定，因为合作代表了双方的长远利益。合作代表双方的长远利益是因为：其一，双方可以通过共同遵守协定而不是共同破坏协定得到更多的好处；其二，从长远的观点看，合作比不合作能给双方带来更多的好处；其三，如果一方违反协定，这将迫使另一方下次也违反协定。这样，行动的后果不仅不符合他们的长远利益，当他们第二次需要合作时，也不符合他们的短期利益。在重复性的“囚徒悖论”情景中，有远见的理性主义者可以不断通过合作与不合作，让急功近利的理性主义者认识到只有合作才能给自己带来真正的即长远的利益，从而将他们转变成为有远见的理性主义者。这样所有行动者的长远利益就可以得到实现，并避免急功近利的理性主义者所无法避免的“次佳化”的后果。

然而，有远见的理性主义者通过合作来实现其长远利益的打算，只有当另一方同是有远见的理性主义者，或者是“可以教育好的”急功近利的理性主义者时，才有可能成功。否则，合作对有远见的理性主义者来说也不是理性的行为。一般说来，合作或遵守协定是一种理性的行为当且仅当如下条件得以满足：

第一，有远见的理性主义者知道另一方和自己同是有远见者，或者属于“可以教育好

的”急功近利的理性主义者，否则，如果另一方是无可救药的、顽固的急功近利的理性主义者，则有远见的理性主义者的“转化”工作所付出的代价就不可能得到补偿，这不符合其利益。

第二，有远见的理性主义者知道对方也知道自己是远见的理性主义者。

第三，有远见的理性主义者知道对方也知道对方是远见的理性主义者。

在一次性“囚徒悖论”情景中，行动者彼此不知道对方要干什么，故上述条件无法满足。在重复性“囚徒悖论”情景中，行动者可以从对方以前的行动中找到对方在这一次可能采取何种行动（合作或不合作）的证据，故上述条件有可能通过反复试探得以实现。但是，如果通过合作和不合作的反复试探，有远见的理性主义者终于发现对方是一个无可救药的“急功近利者”，则有远见的理性主义者为教育和发现无可救药的“急功近利者”所付出的代价就无法得到补偿，但悔之晚矣。换言之，如果对方是一个无可救药的“急功近利者”，即使在重复性的“囚徒悖论”中，合作对有远见的理性主义者来说，也不是一种理性的选择。如何才能保证游戏的另一方是远见者或“可以教育好的”急功近利的理性主义者，从而保证道德行为的充分合理性呢？

#### 四、道德合理性的充分条件

不少哲学家认为，人们需要改变“囚徒悖论”的情景，以保证行动者双方都能遵守协定，为道德的行为提供充分的理由和条件。托马斯·霍布斯最早提出一种政治解决的方案。按照这种方案，为了保证双方遵守协定，双方需要有一个君权或政府（包括警察、监狱、法律制度等强制性的机构）来监督、保证甚至强制双方遵守协定，使得任何违反协定的举止都不可能符合行动者的利益，从而保证实现双方的长远利益。

政治解决的方案确实可以确保游戏各方遵守协议，使“道德的行为”（以及所代表的整体的、长远的利益）和行动者的个人利益完全一致，因而构成行动者采取“道德行动”的充分理由。如果谁要违反道德的要求行事就会受到法律的制裁，其行为就是非理性的。然而，政治解决方案或制度建设只能在法制或监督制度所能有效管辖的范围之内构成代表整体利益的“道德行为”合理性的充分条件，在其有效管辖范围之外则不能。由于经济上的原因，还由于要保护个人的隐私和自由，要建立一个保证任何违反协定者和不道德者都难逃法律制裁的毫无漏洞的制度几乎是不可能的。那么在制度所不能有效管辖的范围之外，在制度的漏洞之内，在类似“囚徒悖论”的情景中，道德的行为和要求还有无充分的理由呢？

有人可能认为整体利益本身就足以提供道德行为的充分理由。但一次性“囚徒悖论”情景的分析证明，如果参加游戏的其他人不合作，不按道德要求办事，只有行动者自己从整体利益出发按照道德要求行事，其结果并不能实现双方的整体利益。有道德的行动者只是牺牲了自己，便宜了那些无限制追求自己个人利益的人。显然在这种情况下，道德的行为并无充分的理由。[5] 日常生活中的例子也能说明这一点。詹姆士（William James）曾经举过一个著名的例子。有两个亡命之徒登上一列火车行凶抢劫。如果车厢里所有的乘客一致起来反抗，则两个亡命之徒只能束手就擒。因为两个人无论多凶，毕竟敌不过一车厢的乘客。然而，如果只有一、二个乘客挺身而出，而其余乘客噤若寒蝉，则挺身而出的乘客必定受到伤害，全车厢人的整体利益也无法实现。其结果使得道德的行为失去了合理的充分性。

那么在重复性的“囚徒悖论”情景中，道德的行为是否有充分的合理性呢？我认为在重复性的“囚徒悖论”情景中，行动者有更充分的理由去做有道德的事情。[6] 如前所述，即使是在重复性的“囚徒悖论”情景中，如果丙某是一个无可救药的“急功近利者”，而丁某并不知道丙某是否可以救药，对丁某而言，遵守协议，按照道德的要求行事，也不符合合理性。但这一情况假设了两个条件：其一，博弈中只有两个行动者；其二，其中一个无可救药的“急功近利者”（即所有其他的交往者都是无可救药的“急功近利者”）。如果博弈中的交往者超过两个，并且并非所有的交往者都是无可救药的“急功近利者”，则按照道德的要求办事优于无限制地追求最大限度的个人利益的行为。[7] 请看如下证明。

让我们在农场主的例子中再加上另一个农场主戊某。假定戊某是一个有远见的理性主义者，他还没有加入丁某和丙某的防守协议。又假定丙某是一个无可救药的“急功近利者”，但丁某是一个有远见的理性主义者。丁某可以通过一到二次合作的代价弄清丙者是一个无可救药的“急功近利者”，从而放弃与丙者的合作，转而寻求新的合作者。在我们设定的情况下，即寻求同戊某的合作。戊某非常愿意同丁某签订防守协议，而不是丙某，因为丁某已经证明自己是一个有远见的理性主义者，而丙某则证明自己是一个不值得信任的、无限制追求自己利益的利己主义者。这样，丁某和戊某就会达成新的防御协定，并且通过相互合作，分享合作的好处。而丙某则被排除在合作之外，无法分享合作的好处。从长远的观点来看，丁某和戊某从相互合作中所获得的好处要远远大于丙某从利己主义的行为中所获得的好处。总的说来，一个人有充足的理由采取道德的行动仅当采取道德的行动所得到的好处（即与有远见的理性主义者打交道所获得的好处）大于他的损失（与急功近利主义者打交道所遭受的损失）。社会中有远见的理性主义者（包括可以转化的“急功近利者”）越多，顽固不化的“急功近利者”越少，一个人就越有充分的理由成为一个有远见的理性主义者（有道德者）。

当然，如果参加博弈的其他人或绝大多数人都是无可救药的“急功近利者”，则任何情况下，行动者都无充分的理由采取道德的行动，尽管采取道德行动的必要理由依然存在[8]。总之，在制度所不能有效管辖的范围之外，在制度的漏洞之内，在道德的行为不能最大限度地扩大个人利益的情况下，道德行为合理性的充分性是一个变量，即对行动者个人而言，是否有充分理由按照道德要求办事取决于参加游戏的其他人是否也按照代表整体利益的要求办，取决于行动者是否相信其他人是否也采取同样的道德的行动，其合理性的充分程度也取决于在游戏中按照道德要求办事人数的多寡。

## 五、合理的道德要求和国民的道德教育

从上面的分析我们可以看到，行动者是否有采取道德行动的充分理由取决于两个基本条件。第一，参加博弈的其他大多数人，或者说社会上的大多数人是否遵守同样的道德规则，采取同样的道德行动。第二，行动者是否相信其他大多数人遵守同样的道德规则，采取同样的道德行动。

代表整体利益的法制和制度建设对于保证实现第一个条件，保证社会大多数人的行为符合整体利益的要求是非常必要的，但却是不够的。原因有三。其一，一个国家不可能对所有代表整体利益的行为作出法律的规定，也没有有效的强制性的手段来保证这种无所不包的法律的有效性。其二，即使在法制管辖的范围之内，如果一个人并不是从整体利益出发自觉遵守，而是从个人利益出发不得不遵守，他就会想法设法钻法律和制度的漏洞，以最大限度地牟取个人利益。这样就会扩大制度的漏洞，减低制度的效率，加大制度建设的成本。其三，如果整个社会都充满了这样的利己主义者，都来寻找制度的漏洞，行贿受贿，这就会造成法不责众的情况，就会出现有法不依，执法不严，违法不办的情况。所谓制度就会徒具形式。因此，为了保证第一个条件，我们需要培养自觉按照整体利益要求行事的有道德素养的人，以增强制度的效率，弥补制度的不足。

美国哲学家科夫卡(Gregory Kavka)认为有两种约束力影响人们的道德行为。一种是外在的约束力(external sanctions)，即指警察、监狱、法律制度等。另一种影响人们行为的力量是内在的约束力(internal sanctions)，即人们的道德感、良心、或良知。科夫卡将其定义为“赏善罚恶的心理结构”。[9] 培养有道德的人就是要培养有“赏善罚恶的心理结构”的人。他们在做不道德的事情时会感到良心上的不安和内疚，这种良心上的不安和内疚会约束他们做不道德的事情。反之，做道德的事情可以满足他们的道德感和良心，这种满足感会鼓励他们做有道德的事情。由于人们的良知内在于心，故无所不在，因而比外在的约束力更有效，同时也不会侵犯个人的隐私和个人的自由。

为了保证有道德人行为合理的充分性，我们需要有更多的、更自觉的、有“赏善罚恶心理结构”的人，需要有在制度的有效管辖范围之外还能从整体利益出发采取行动的人。否则，整个社会充斥只考虑个人利益，完全不考虑整体利益的人，再完备的法制，也无法使制度有效管辖范围之外，制度漏洞之内的道德行为有充分的理由，在那种情况下，道德之士只能成为利

己之人所利用的牺牲品。

关于行动者有无充分理由采取道德行动的第二个条件和认识的理由有关，这是构成道德行动的充分理由的非常重要的方面。美国“九·一一”事件中，被劫持的飞机上不乏有理性的道德之士，其中三架飞机上的乘客没有采取反抗行动，一个重要的原因是缺少认识上的充分的理由，主要有两个：一是不知其他乘客是否会采取同样的行动（因为如果其他乘客不动，则反抗不会有效，徒做无谓牺牲）。二是不知道这次劫机后果的严重性。这和“囚徒悖论”情景相似，行动者如果缺少充分的认识理由相信他人会合作，他就没有充分的理由采取合作的行动。“九·一一”之后，人们对恐怖主义分子劫机后果的认识加深，当有一位恐怖主义分子企图在飞机上将藏在鞋底的炸弹点燃时，周围的乘客几乎是一拥而上，将其制服，因为从认识上，人们有充分的理由相信其他人一定会采取同样的行动。

怎样创造第二个条件？我认为需要创造良好的社会风气和互信。一个人究竟有无理由相信其他人也会采取同样的道德行动取决于整个社会风气和互信。我从报上曾读过两则消息。一条消息说，一个傍晚，在一个小镇的一条街上，有一位年轻的妇女被一位歹徒追打致死，现场惨不忍睹。年轻妇女凄厉的呼救声响彻小镇的上空，整个过程长达一小时，听到和看到这一切的不止一个人，但没有一个人出来阻止，甚至喝止。也许，如同那位记者所说，“一条街的良知都冻死了。”但也不可否认，他们当中也许有人想站出来，但不知道其他人是否能一起站出来有效地制止这种残暴的行为，从而退缩。另一条消息则说，一个深夜，一位歹徒将一位妇女打得头破血流，将其钱包抢走，但一位男子听到呼救声，拿起家伙就下楼追赶，住宅小区凡听到的人几乎不约而同跑出来追赶歹徒，歹徒如过街老鼠，很快就被抓获。有了这样的社会风气，人们才有充分的理由相信在做有道德的事情时，尤其是见义勇为时，其他人也会采取同样的行动。

怎样创造良好的社会风气？除了制度建设以限制损人利己的行为，以保护有远见的理性主义者（即讲道德者）的利益外，我们还需要提高整个国民的道德素养和道德品质。怎样提高整个国民的道德素养？我想除了要从小抓起外，从道德的内容上来说，我们要讲求道德要求的合理性。首先，它要有可能代表整体利益。[10] 其次，它合理的充分性和讲道德人的个人利益成正比，即越符合讲道德者的利益，对其利益损害越小，其合理程度越充分。反之，充分性则降低。道德要求合理性越充分，就越能为大多数人所自愿接受并实践。[11]

根据上面提出的两条原则，我想提出一种“最低限度”的道德要求的概念。其基本内容主要包括不要“损人利己”和对自己个人利益损害、影响不大的情况下的利他要求（如不要随地吐痰，不要乱仍果皮纸屑，等）。这是一种尽可能考虑到行动者个人利益的利他要求。这是一种对无限制扩大个人利益行为进行限制和约束的道德要求。至于在特殊的情况下，如战争，为了国家和整体的利益，要求个人做出重大牺牲，甚至生命的利他要求，则可以通过某种“契约”或法律的形式去实现，而不是仅仅依靠道德的约束。

这种“最低限度”道德的合理的必要性是显而易见的，因为她代表了整体的利益。而其合理性的充分性也是非常明显的，因为她尽可能地保障了讲道德人的利益。这种最低限度的道德要求甚至不排除见义勇为的行为——如果见义勇为发生在良好社会风气的社会中，发生在大多数人都能挺身而出的情况下，这样的见义勇为甚至也是这种最低道德要求所要求的。对于有可能牺牲生命的见义勇为，在正常情况下，这应该是警察的职责，对做不到这一点的普通国民，道德上不作谴责（对能这样做的普通国民，道德上当当然要加以肯定。这就像从法律上我们规定私有财产不可侵犯，但我们不能因此认为自愿将其部分私有财产捐给社会的人是违法的）。和过去那种似乎时时刻刻都要求普通人们“毫不利己，专门利人”的道德要求相比，这种“最低限度”的道德要求的合理性是显而易见的。现在很多人不光是干“损人利己”的事情，而且还干“损人不利己”，“损人损己”的事情，从实践理性的角度，这也是非理性的，这是对非理性的利他主义的一种逆反。“最低限度”的道德要求以其合理性可以逐渐克服这种非理性的不道德的行为，最终会为大多数人所接受并实践，从而促成良好的社会风气的形成，而良好的社会风气和互信的形成又会为道德行动的合理性（甚至是见义勇为式的道德行为的合理性）奠定基础。

## 参考文献和注释：

- [1] 关于西方哲学家的这种普遍的看法，参见David Gauthier, “Morality and Advantage,” *Philosophical Review* 76 (1967): pp.460-475, Thomas Nagel, *The Possibility of Altruism* (London: Oxford University Press, 1970), p.3, David Gauthier, *Morals By Agreement* (New York: Oxford University Press, 1986), p.1, 和 Jean Hampton, “The Wisdom of the Egoist: The Moral and Political Implications of Valuing the Self,” *Social Philosophy & Policy*, Vol. 14, No. 1 (Winter 1997): pp.21-22。
- [2] 比方说，有人认为整体利益就包含了个人的利益，这种说法或多或少地混淆了两者之间的区别。“囚徒悖论”的情景将这种区别表达得非常清楚。
- [3] 有的西方哲学家认为还有一种既不依赖自己也不依赖他人的纯客观的价值，本文不打算讨论这个问题。
- [4] 上述例子源于Ishtiyaque Haji。参见Ishtiyaque Haji, “Hampton on Hobbes on State-of-Nature Cooperation,” *Philosophy and Phenomenological Research* Vol.LI, No.3 (September 1991): 594-595。
- [5] 但美国哲学家哥梯尔认为即使在一次性的“囚徒悖论”情景中，行动者仍然有充分的理由遵循道德的原则。关于哥梯尔的论证，参见 David Gauthier, “Morality, Rational Choice, and Semantic Representation: A Reply to My Critics,” *Social Philosophy & Policy*, Vol. 5, No. 2 (1987)。关于对哥梯尔论证的批评，参见拙文“哥梯尔的‘协议道德’理论简评”，《河北学刊》2004年第3期。
- [6] 关于这一点的证明，可参见David Gauthier, *Morals by Agreement* (Oxford: Clarendon Press, 1986), 157-189。
- [7] 这一点已为某些计算机程序所证明。参见Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984) 和 Matt Ridley, *The Origins of Virtue* (New York: Viking Penguin, 1997), 60-61。
- [8] 关于道德行为的必要理由和条件，见本文第三部分。
- [9] 见 Gregory Kavka, “The Reconciliation Project” in *Morality, Reason and Truth* ed. D. Copp and D. Zimmerman (Rowman and Allanheld, 1984), 305。
- [10] 我用“有可能代表整体利益”，因为道德行动的结果并非总是反映了整体的利益。比方说，“囚徒悖论”中，合作（即道德的行为）是代表整体利益的行为，但能否实现整体的利益取决于是否所有的行动者都采取合作的行为。从这个意义上说，合作并非必然代表整体的利益。因此，具体的道德行为是否真的能实现整体利益，有多大可能实现整体利益，也是判定道德充分合理性的依据之一。
- [11] 有人也许会提出疑问：道德行为要尽可能地符合个人利益，那么道德的要求和利己主义岂不是没有区别？两者之间当然是有区别的。首先，道德要求行动者从整体利益出发采取行动，但利己主义要求行动者仅仅从个人利益出发。其次，在类似“囚徒悖论”的情景中，在行动者知道参加博弈的其他人将遵守道德规则的情况下，那么，即使违反道德规则会更有利于行动者，道德也要求行动者遵守道德规则。然而，利己主义则要求行动者违反道德规则，按照自己的利益行事。

（本文发表在《云南大学学报·社会科学版》2004年第6期。作者现将“囚徒悖论”译为“囚徒困境”。）

作者电子信箱: [chenzhen@nju.edu.cn](mailto:chenzhen@nju.edu.cn)

/