

# 制度如何激发社会合作

## ——基于公共品博弈实验的前沿理论综述

章平<sup>1</sup> 许志成<sup>2</sup> 闫佳<sup>1</sup>

(1. 深圳大学中国经济特区研究中心, 广东 深圳 518060; 2. 美国德克萨斯州农工大学经济学系, Texas 77843)

**【摘要】** 公共品博弈作为刻画人类社会协调与合作这两个问题的基本模型, 成为经济学、管理学、政治学、人类学等跨学科的研究热点和难点。通过分析引入激励制度的公共品博弈实验研究, 比较奖和罚两种制度如何改变人们的信念并影响公共品合作供给行为。将已有公共品博弈实验研究分为外部激励和内部激励(决定赏罚规则、强度和执行方式的制度)、货币激励和非货币激励(道德劝说、训斥和社会排斥), 考察不同制度对于提升公共品合作供给水平的实际作用, 特别是惩罚制度的社会净效率。在此基础上, 述评已有研究成果的一般性结论和启示, 展望有关个体偏好、社会合作和制度发育的理论和实证研究可能的创新。

**【关键词】** 公共品博弈; 合作; 奖励; 惩罚; 内生制度

**【DOI 编码】** 10.13962/j.cnki.37-1486/f.2015.01.004

**【中图分类号】**F224.32 **【文献标识码】**A **【文章编号】**2095-3410(2015)01-0026-08

在涉及集体行动问题时, 作为该问题的核心, 如何保证公共品的有效供给依然没有明确的结论。根据经典博弈论对个人行为策略的预测, 公共品通常不能经由自愿供给机制(下称 VCM)来得到有效的提供。<sup>①</sup>这是因为建立在理性经济人假设基础上的标准非合作博弈理论认为, 搭便车行为将成为参与者的占优策略, 那么在均衡状态下将自然没有人自愿供给公共品。然而, 现实情形下的大量事实却显示, 即使在 VCM 中, 个体行为也不曾像实验室和田野研究的悲观预期那样。究其原因, 则是因为隐藏在集体生活中的个体自发合作并不仅仅受利己偏好所驱动。虽然有学者已经总结出成功合作模式所具有的共同点, 但是令人遗憾的是 VCM 中的供给水平会随时间而衰减, 如果没有进一步的激励, 合作将有瓦解可能。因此, 经济学家们一直都在寻找更好的方法来解决公共品的低效供给状态。一种方法是使

用外部激励, 另一种方法则是引入内部激励机制。受 Fehr 和 Gächter (2000)<sup>[1]</sup> 的启发, 公共品实验中已经出现了大量关于内部激励制度的研究, 既包括对货币奖励和惩罚的比较, 也包括非货币惩罚。

### 一、公共品博弈实验的理论缘起

(一) 自愿供给机制下的一般线性公共品博弈  
自愿供给机制作为一个一般的线性公共品博弈不存在任何的沟通或是激励, 典型的自愿供给机制公共品博弈实验设置如下: 组内有  $n$  个参与者, 在重复公共品博弈的每一轮每一个参与者都被给定初始禀赋  $w$  来决定如何分配自己的初始禀赋。如, 分配给自己私人账户的有多少, 分配给公共账户的有多少。那么, 对于参与者  $i$  来说, 其在每一期的收益为:

$$\pi_i = (w_i - c_i) + \alpha \sum_{j=1}^n c_j \tag{1}$$

其中,  $c_i$  为参与者  $i$  对于公共品的贡献,  $\alpha$  为公

**【基金项目】** 本文是国家自然科学基金项目“公共品自愿供给中的社会偏好与惩罚机制研究”(项目编号:71203145)和深圳市哲学社会科学“十二五”规划课题“深圳城中村共享资源自治理机制研究”(项目编号:125A134)的阶段性成果。

**【作者简介】** 章平(1981-), 男, 浙江绍兴人, 深圳大学中国经济特区研究中心副研究员, 管理学博士。主要研究方向: 制度经济学、行为经济学。



共品的边际收益 (MPCR), 且  $0 < \alpha < 1 < n\alpha$ 。这一不等式说明, 在其存在的条件下, 尽管对于公共品的完全供给为社会最优, 但也存在 0 供给这一纳什均衡。在许多的实验设置中,  $w = 10, \alpha = 0.4, n = 4$ 。在  $t$  期, 参与者  $i$  将 5 禀赋分配到自己的私人账户, 将 5 禀赋作为对公共品的供给。与此同时, 其他三个参与者的总供给量为 20。那么, 参与者  $i$  在  $t$  期的收益就为 15。对于参与者  $i$  来说, 会非常容易意识到对于公共品供给的占优策略为 0。在此例中, 参与者  $i$  将 10 禀赋放入自己的私人账户, 并会得到 18 个单位的收益。因此, 根据纳什均衡所预测的结果, 理性自私的参与者将不会有动力去供给公共品。然而, 实验结果却显著不支持纳什均衡所预测的结果。<sup>②</sup> (1) 在一次性博弈和有限重复博弈中的首轮, 被试者愿意将自己财富的 40% - 60% 供给投入到公共品中。(2) 首轮之后的供给水平趋于降低, 但仍大于零, 而末轮博弈中, 超过 70% 的被试者选择零供给。尽管如此, 若不存在进一步的激励机制, 合作将不会维持。因此, 设计一个有效的激励机制来提高公共品的供给水平依然非常必要。

(二) 将惩罚引入公共品博弈的经典文献

Fehr 和 Gächter (2000)<sup>[1]</sup> 研究了对搭便车者的货币惩罚。其中, 博弈的每一期包含两个阶段: 第一阶段, 被试者同时在私人账户和公共账户间分配自己的禀赋, 然后获取收益和组群内每一个其他被试者供给水平的信息反馈; 第二阶段, 所有被试者决定对于其他被试者有成本的惩罚。因为存在惩罚, 则对于组群内的每一个被试者  $i$  其收益为:

$$\pi_i = [(w_i - c_i) + \alpha \sum_{j=1}^n c_j] \left(1 - \frac{1}{10} P^i\right) - \sum_{j \neq i} c(p_j^i) \quad (2)$$

其中,  $p_j^i$  为被试者  $i$  对被试者  $j$  的惩罚, 且  $P^i = \min(\sum_{j \neq i} p_j^i, 10)$ ,  $c(\cdot)$  为惩罚的成本函数, 递增且凸。

在第二期, 惩罚别人则为非理性, 如在理性假设下惩罚威胁的不可置信。根据逆向归纳法, 对于每一个参与者来说, 子博弈精炼均衡依然为 0。然而, 结果显示惩罚可能性的存在提高了公共品的平均供给水平, 并且无论惩罚引入的顺序或者固定组群效应, 或者博弈中的随机匹配, 惩罚均可以显著提高公共品供给的合作水平, 而且具有稳健性。

为什么参与者会选择惩罚他人? 社会偏好可以很好地解释这个问题。许多实验研究讨论了到底是什么因素在驱动参与者在公共品博弈中对搭便车者进行惩罚, 如对等性和对于公平的关心等。Casari 和 Luini (2012)<sup>[2]</sup> 的实验分析显示, 从事于合作任务的被试者并不是因为策略原因, 而是因为情感动机才进行合作的。换句话说, 他们并不将惩罚视作提高合作水平的策略, 而是将之视为对于公平关心的情感表达。

(三) 经典文献的不足

Fehr 和 Gächter (2000)<sup>[1]</sup> 的研究仅仅为公共品博弈中同侪惩罚机制的一个开始, 之后出现了大量文献, 但是一些问题依然悬而未决。

第一, 惩罚机制的有效性依然模棱两可。如果惩罚, 或是有惩罚的意愿作为一个不可置信的威胁, 则有效供给水平不能长期维持。换句话说, 只要搭便车行为真的被惩罚了, 惩罚确实可以阻止搭便车。然而, 惩罚存在社会成本, 因此对于社会效率的考量会使得惩罚的有效性下降。基于此, 我们依然需要寻找其他的一些机制来提供不同的激励。

第二, 在现实生活中, 对搭便车者的惩罚可能遭致搭便车者的报复等反惩罚行为, 这就使得人们打消了惩罚搭便车者的念头。Nikiforakis (2008)<sup>[3]</sup> 在公共品博弈实验中引入反惩罚机制, 并比较了自愿供给机制、Fehr 和 Gächter (2000)<sup>[1]</sup> 的惩罚机制和反惩罚机制所带来的效率。反惩罚机制几乎与惩罚机制一样, 仅仅多了一个阶段 (如第三阶段), 在这一阶段, 所有被惩罚者可能会报复那些在第二阶段惩罚他们的被试者。由于存在报复或者说反惩罚的威胁, 惩罚搭便车者的意愿随之降低。结果, 反惩罚机制破坏了公共品博弈中促进合作提升的制裁机制的效力。在一些不发达国家, 反惩罚机制极其严重。Herrmann et al. (2008)<sup>[4]</sup> 的田野实验比较了世界范围内 16 个带有反惩罚机制的公共物品投资项目, 研究发现反社会惩罚强大到足以破坏内部激励制度所产生的合作增强效应。并且, 研究将反惩罚行为与公民合作规范、法律系统脆弱性联系起来。

第三, 由于所谓的“挤出 (crowding out)”效应使得货币惩罚的效力变得更加模糊。<sup>③</sup> 如果忽视“隐藏成本”, 惩罚机制的效力可能会被高估。因此, 施行



货币激励有时可能对合作和信任带来负面影响 (Bowles 和 Hwang, 2008)<sup>[5]</sup>。相反,道义上的规劝作为一种非货币激励形式的表现却不会比惩罚差 (Masclet et al., 2003)<sup>[6]</sup>。

最后一点,也是最重要的一点,近年来开始出现了一股新的潮流,即允许激励制度安排的民主式参与。与大多数类似 Fehr 和 Gächter (2000)<sup>[1]</sup>形式的研究所采取由实验者外部施加惩罚机制不同,内部的制度设置允许被试者自己选择他们所偏好的激励机制。

## 二、公共品博弈中的赏罚制度

为了追求公平,人们愿意为惩罚搭便车者而牺牲自身福利。在自然田野实验中,个体也使用奖励来激励团队生产和公共品博弈中的合作行为。因此,在实验室实验中引入一种新的激励机制则是内部激励制度的自然发展。与惩罚一样,奖赏也与自利效用最大化假说不一致。奖励与正对等性有关,不同于惩罚(负对等性)所起的作用方向,奖励的一个显而易见的优势是它不会像惩罚那样遭致社会效率的无谓损失,但在实验室实验中它还会发挥这种在自然田野实验中所发挥的作用吗?

### (一)奖励如何促进合作

关于社会偏好的文献对正对等性和负对等性的动机与差异进行了深入分析,但是这些文献的目标并不是比较奖励和惩罚在公共品供给情境下的效力差异,这是因为对等性与公平的考量对于博弈情境和过程相当敏感,因此洞悉公共品博弈中奖励的作用如何不同于惩罚则非常必要。

Andreoni et al. (2003)<sup>[7]</sup>的实验设计使用了一个简单的“提议—响应博弈”,首次对奖励和惩罚进行了比较。他们发现响应者倾向于牺牲自己的所得来奖励更加慷慨的提议者,惩罚更自私的提议者。在一个典型的“提议—响应博弈”中,Andreoni et al. (2003)<sup>[7]</sup>分析了四种不同的处理方式,分别是独裁者、胡萝卜—大棒(惩罚—奖赏)、胡萝卜(奖赏)、大棒(惩罚)。在每种处理的第一阶段,一个提议者分配 2.40 美金给响应者。这四种处理方式唯一不同的地方就在于第二阶段响应者是否有能力可以惩罚提议者。在独裁者处理中,响应者既不能惩罚也不能奖励,这与公共品博弈中的自愿供给机制相一致。

大棒处理中,响应者可以用 1 美分的成本来惩罚提议者,使之收益减少 5 美分,这与 Fehr 和 Gächter (2000)<sup>[1]</sup>中的惩罚相似。在胡萝卜处理中,响应者可以用 1 美分的成本来奖励提议者,使之收益增加 5 美分,这与公共品博弈中的奖励机制相似。此外,在胡萝卜—大棒处理中,很多发现都与常识相一致,给出了奖励和惩罚可以共同存在的证据。并且,在胡萝卜—大棒组的平均提议最高,独裁者博弈组的平均提议最低。因此,奖励和惩罚可以共同起好的作用。一个惊喜的发现则是,相比于惩罚,奖赏可以带来更高的合作水平。

尽管有很多相似的地方,但提议—响应博弈却并不能包括团队生产的情境。另外,奖励也并不是纯粹的再分配,而是一个更强的正对等性。因此奖励促进合作的效力可能被高估。

### (二)奖励或惩罚在公共品实验中的表现

在公共品供给实验中,是否奖励对于促进合作要优于惩罚? 大多数文献认为,惩罚机制会维持高水平的贡献。然而,如果惩罚可置信且如实发生,那么合作就会保持在较高的水平,由于惩罚会带来无谓的损失,所以平均收益的比较依然不明朗。

受 Andreoni et al. (2003)<sup>[7]</sup>启发,Sefton et al. (2007)<sup>[8]</sup>将前者的实验环境扩展到公共品供给实验中,并直接比较惩罚和奖励在维持公共品供给的效力。在控制组,含有四个成员的所有组别都进行一个 10 轮公共品自愿供给序贯博弈。10 轮之后,内部激励制度被引入。和 Andreoni et al. (2003)<sup>[7]</sup>相同,每一轮都有两个阶段,被试在第一阶段决定自己的贡献水平,在第二阶段除了控制组继续进行和前 10 轮一样的公共品自愿供给序贯博弈外,取决于处理方式的不同,其他被试者在自己的处理组别内分别被赋予奖励(胡萝卜)、惩罚(大棒)、或者奖励加惩罚(胡萝卜—大棒)的权力。需要强调的是,与 Andreoni et al. (2003)<sup>[7]</sup>的一个重要的不同,Andreoni et al. (2003)<sup>[7]</sup>的奖励是一个很大数量的净效率所得(强正等对性),因为一个单位成本的奖励会提供 5 个单位的收益所得,而 Sefton et al. (2007)<sup>[8]</sup>的奖励仅仅只是一个纯粹的收益再分配。

前述实验给出的结论显示,在没有惩罚和奖励的控制组,公共品的贡献和收益随博弈的重复进行



而恶化。与 Andreoni et al. (2003)<sup>[7]</sup> 相比,公共品自愿供给博弈要比提议一响应博弈更复杂。因为提议一响应博弈仅仅只是两个参与者一对一的相互作用,在公共品自愿供给博弈中,个体在不同的组别会对激励做出不同的考虑和反应。在奖励组别,初始阶段的贡献水平随着响应者的奖赏激励而增加。然而,之后的贡献水平稳步跌落,甚至跌到没有奖励的组别贡献水平之下。因此,单独的奖励并不能维持贡献水平。甚至更糟的是,当合作者逐渐发现奖励并不能减少搭便车行为时,利他动机则会受到严重伤害。相反,惩罚可以维持平均的贡献水平,但惩罚可能会带来无谓的社会效率损失,而胡萝卜加大棒的处理方式所带来的协同作用得到了最高的贡献与收益。

然而,惩罚在促进合作上要优于奖励的观点还是受到了质疑(Rand et al., 2009)<sup>[9]</sup>。Fehr 和 Gächter (2000)<sup>[1]</sup> 的实验中,为了移除可能混淆惩罚影响的声誉,组内成员在每一期都采取匿名互动。但使用匿名制奖励实验可能被误导,因为在真实的生活中,人们之间的相互关系并不完全是匿名的。没有匿名制的话,由于声誉和对社会影响的顾虑,个体间紧密的社会关系将改变他们的行为。<sup>④</sup> 因此,Rand 等(2009)<sup>[9]</sup> 认为 Fehr 和 Gächter (2000)<sup>[1]</sup> 仅仅考虑惩罚的影响,而没有考虑声誉影响也是合意的,但当我们要比较惩罚和奖励时,这又会产生误导,因为在此环境下人们不能建立起长期对等性的关系。因此,即使个体明白奖励将有助于建立长期的对等性,他们也没有在前面所提及的实验环境中使用它。

Rand et al. (2009)<sup>[9]</sup> 的实验设置几乎与 Sefton et al. (2007)<sup>[8]</sup> 一致,不同的是,Rand et al. (2009)<sup>[9]</sup> 给被试者分配固定身份,这实际上允许定向互动和直接的对等性(紧密的社会关系),Sefton et al. (2007)<sup>[8]</sup> 的实验设置则给被试者随机分配身份,这实际上为匿名互动和间接的对等性。在定向互动和直接的对等性下,他们发现奖励和惩罚一样都可以维持合作。由于惩罚存在无谓的效率损失,奖励、奖励—惩罚两种处理方式所带来的平均收益均高于惩罚处理方式。据此,他们认为在重复博弈的公共品供给实验中,惩罚对于维持合作依然是必

要的,相反,奖励则会提供一个更好的效率互动。

### 三、公共品博弈中的内生激励制度

前述的公共品博弈实验中激励机制都是由实验者所外生给定,近年来研究者在实验室试验中开始探究通过民主协商、内生发育的激励机制是否与外生给定不同。很多的实证研究都认为民主参与能促进合作行为,例如 Bardhan (2000)<sup>[10]</sup> 的实证研究发现,在印度当灌溉规则是由当地农民自己制定的,那么他们就很少不遵守这一灌溉规则,这是因为决策过程可以影响人们对公平及个人之间如何合作的态度,即人们不仅关心分配的结果,而且也关心分配的程序。

Dal B et al. (2010)<sup>[11]</sup> 所进行的一个实验发现,当激励机制根据被试者之间的民主协商来选择,那么所带来的合作水平高于外生给定激励机制的合作水平。尽管他们使用的是囚徒困境博弈,而不是公共品博弈。Kosfeld et al. (2009)<sup>[12]</sup> 设计了一个公共品博弈实验,在实验中参与者可以选择这样一种制度:每个被试者如果决定加入公共品博弈,他将和其他被试者一起参与博弈实验;如果决定不加入公共品博弈,则将实验前分配的初始禀赋全部放入自己的私人账户。这一制度形成后,当且仅当所有被试者同意,这项制度才能被执行。制度一旦被执行,任何搭便车行为都将会被外部权威惩罚。实验提供了自治提高合作和效率的证据,但是这一实验并没有考虑同侪惩罚机制。因此,将民主参与引入实验研究内生激励机制的有效性,需要在公共品博弈中设计一个可行和有效的激励机制。

#### (一) 惩罚强度的决定以及惩罚的实施

Walker et al. (2000)<sup>[13]</sup> 首次将投票表决引入实验室公共资源博弈中,组内成员可以由大多数或全体一致的原则来表决供给的最低水平。当供给低于最低水平时,由第三方来进行惩罚,结果发现表决提高了效率。Kroll et al. (2007)<sup>[14]</sup> 将研究扩展到如 Fehr 和 Gächter (2000)<sup>[1]</sup> 一样的同侪惩罚,结果认为表决自身并不能提高合作,但是如果表决者可以惩罚违规者,供给贡献就会提升。并且,尽管有成本的惩罚会牺牲社会效率,表决群体的总效率还是要高于没有表决机制群体的效率。Ertan et al. (2009)<sup>[15]</sup> 允许参与者对惩罚制度进行表决,即在没



有惩罚、有约束的惩罚、中等程度的惩罚以防止滥用惩罚中选择。他们发现被试者从开始阶段不允许任何惩罚,逐渐过渡到允许一个清晰的对于低供给者的惩罚,但是没有一组同意不受约束的惩罚制度。

尽管一些研究比较了惩罚和奖励的效力,但依然还存在一些问题,如参与者偏好哪一种制度?如果制度由参与者内生决定而非外部实验者施加,将会怎样? Ertan et al. (2009)<sup>[15]</sup> 将是否执行没有内生激励制度的标准 VCM、奖励或者惩罚的 VCM 的权力授予给同一组内的被试者。在开始进行重复公共品博弈前,组内成员就激励机制进行表决。可供选择的包括内生奖励、内生惩罚和标准 VCM。他们发现内生激励并不是非常有效,或者从技术上来讲,惩罚成本和被惩罚成本比为 1:1,标准的 VCM 将成为首选。但是如果内生激励足够有效,如成本比为 1:3,实验结果显示在提高供给水平时,惩罚是最有效的。

比较 Rand et al. (2009)<sup>[9]</sup> 和 Sutter et al. (2010)<sup>[16]</sup> 的研究,可以得到几点启发性的结论:两者的被试者都被分配固定的身份,并且每一个被试者在每一期同一组内大家都知道其身份,所以被试者之间将会建立声誉。但在每一期的第二阶段, Rand et al. (2009)<sup>[9]</sup> 提供了详细的信息,包括执行惩罚或者奖励的被试者身份,而 Sutter et al. (2010)<sup>[16]</sup> 选择不公开执行惩罚或者奖励的被试者身份以避免发生报复。这两者并不是一个镜像,因为后者将高估惩罚的效力,低估奖励的效力,而前者正好相反。另一个主要的不同是 Sutter et al. (2010)<sup>[16]</sup> 中的激励机制是由组内成员内生决定,他们认为惩罚制度效果较弱的主要原因是人们不愿意服从由实验者外生强加的制度。然而,当参与者可以通过表决来选择激励机制,自治制度将更可能达到合意的效果。

## (二)独裁者对惩罚和奖励制度的偏好

激励机制的内生安排并不必然表示为民主参与的形式。有时还可以采取这样一种激励方式,即惩罚权被集中到组群内少数人手中,由领导者或者独裁者来决定赏罚。Gürek et al. (2009)<sup>[17]</sup> 尝试回答当惩罚和奖励在一个团队都可供领导者使用时,领导者到底是偏好使用奖励还是惩罚来促进合作。

在一个团队内,领导者被随机选择,并给定一定数量的额外资金用于分配给组内其他成员的惩罚或者奖励,也可以将剩下的资金放入自己的私人账户。他们发现:第一,当执行惩罚时,总合作和效率较高;第二,尽管大多数(95%)领导者在初始阶段选择奖励制度,但选择惩罚制度的领导者比例却会一直增高,因为他们逐渐发现惩罚可以更好地维持供给;第三,合作的路径依赖,如不同激励机制使用的顺序对合作和效率相当重要。当然,这些发现与政治经济学的许多研究相悖,政治经济学中的独裁者并不是仁慈的神,而是倾向实施惩罚的人。

## 四、非货币激励制度

如前所述,非货币惩罚不会在明确的环境下带来效率损失,并且会有效促进合作行为。本部分将总结非货币激励的不同方式。

### (一)道德劝说和训斥

Masclet et al. (2003)<sup>[6]</sup> 给参与者提供了表达对于他人贡献不满的机会。表达不满不是一种标准的惩罚机制,因为它不能对目标参与者施加直接的货币成本。于是,根据标准的基于自利偏好的理论,道德劝说或训斥不能改变被试者的行为。然而,实验结果给出了违反经典理论的结论。研究者比较了非货币惩罚和货币惩罚的效力,他们发现与基准的没有惩罚机会的 VCM 相比,两种惩罚机制都能提升贡献水平。虽然货币惩罚中的平均贡献水平较高,但是货币惩罚却会遭致效率损失,结果非货币惩罚中的平均收益甚至更高。

Noussair 和 Tucker (2005)<sup>[18]</sup> 的实验中货币和非货币惩罚都可用,从而扩展了 Masclet et al. (2003)<sup>[6]</sup> 的工作。他们发现,尽管非货币惩罚在一开始可以提升贡献水平,但是非货币惩罚中的合作还是随实验的继续进行而下降。因此,道德劝说仅仅在有货币惩罚作为支持时才是有效的。然而,道德训斥这种负向非货币激励并没有表现出稳健的有效性。Hopfensitz 和 Reuben (2009)<sup>[19]</sup> 检验了社会困境实验中情绪表达的有效性,他们发现情绪表达不足以提升合作水平。实际上,愤怒也会激发受罚者的报复,削弱情绪机制的有效性。与此同时,像内疚一样的社会情感被认为是对不合作行为的一个有效的阻碍。因此,对于情感机制通过何种渠道可以



提升合作水平,仍有待深入研究。

## (二)公共品博弈中的社会排斥

另一种非货币惩罚就是所谓的排斥,排斥为组群成员通过投票规则从组群内驱逐一些人。排斥的有效性与组群成员的异质性有关,一些开创性的实验研究发现了异质性在公共品博弈中的重要性(Fischbacher 和 Gächter, 2010)<sup>[20]</sup>。他们认为大多数的参与者是条件合作者,被试者根据自己对于其他参与者平均贡献的信念来调整自己的贡献水平。乐观主义者相信其他人将在高水平上作出贡献,所以他们也在同样的水平上做出贡献(如社会匹配)。但是久而久之,因为逐渐发现了搭便车者,乐观主义者被这种悲观情绪传染而降低贡献。换句话说,悲观信念在组群成员中传染。因此,通过排序分类提高个体信念和偏好同质程度会提升合作水平。

一种方法是外生排序,如 Page et al. (2005)<sup>[21]</sup>的排序方法。在每一期,他们将 16 个被试者分成 4 组进行公共品博弈(VCM),然后他们被按照贡献水平重新分组。四组中最高的贡献者被分到一组,以此类推,接着公共品博弈重新开始。外生排序可以提高合作水平,但是在后期则无效。

Mcevoy et al. (2011)<sup>[22]</sup>设计了一项内生排序机制,其中组内成员被允许通过多数投票规则驱逐搭便车者。在每一期的开始,16 位参与者在搭档协议下被分入一组,然后开始进行每一轮两阶段的 15 轮公共品博弈。在第一阶段,贡献和收益被确定,组员通过多数投票规则驱逐一位参与者。作为一种惩罚,被驱逐的参与者将以禀赋的一半进行同样的公共品博弈。尽管存在多数投票规则,驱逐却并不经常发生,但驱逐的威胁也足以防止搭便车行为。

此外,Charness 和 Chun - Lei (2008)<sup>[23]</sup>扩展的驱逐机制中,通过投票,参与者既可以选择退出也可以选择驱逐其他组员,这一内生的排序制度可以提升合作水平,而且在后期依然非常有效。Maier - Rigaud et al. (2010)<sup>[24]</sup>发现,引入驱逐后,除第一和最后一期,贡献水平显著提升。由于公共品的性质,参与者需要权衡组群规模与平均收益之间的正向联系,异质性组群规模是合作的障碍。由于驱逐,组群规模降低,但平均收益的净效益依然为正。

## (三)用脚投票

Gürek et al. (2006)<sup>[25]</sup>授予被试者自由选择加入没有惩罚的体制或者存在惩罚的体制,如用脚投票(自愿加入或离开公共品供给团队)。在 30 轮博弈中,每一轮的第一阶段,被试者有机会选择加入没有惩罚或者有惩罚的体制中。在第二阶段,参与者进行一个标准的线性公共品博弈。对于那些加入惩罚体制的被试来说,他们可以惩罚任何人。惩罚者 1 单位的惩罚成本带给受罚者 3 单位的惩罚成本,而奖励则是纯粹的再分配。在初始阶段,被试者偏好无惩罚体制,但在后期,当他们意识到惩罚制度可以带来更高水平的贡献和收益,他们将逐渐转移到惩罚体制中。

与前述实验结果一致,参与者在博弈开始时不愿意惩罚搭便车,然而,惩罚和无惩罚体制的平均贡献差距依然非常显著。因此,理性的参与者将转向惩罚体制以得到更高的预期收益。虽然如此,在惩罚体制和无惩罚体制之间的流动不能有助于降低异质性,这对合作来说是一个阻碍。实验的一个自然扩展是允许参与者有可替代的选择,强度、社会距离和激励体系可以带来潜在的制度安排,更多的制度选择可能有助于降低异质性和进一步地促进合作。

## 五、总结

本文回顾了自 Fehr 和 Gächter (2000)<sup>[1]</sup>以来的内部激励制度,因为在没有激励机制的标准公共品博弈中,合作将随时间而恶化。在公共品博弈中经济学家一直在寻找提升合作水平的机制,Fehr 和 Gächter (2000)<sup>[1]</sup>提出惩罚可以促进高水平的贡献。尽管子博弈精炼均衡在完全理性假设下对每一个参与者来说依然为 0,他们的研究认为惩罚机会的存在提升平均贡献水平。然而,惩罚制度的社会净效率尚无定论,因此找寻另一些方法还是非常必要的。

除了惩罚,人们也愿意在团队生产和公共品博弈中使用奖励。奖励比惩罚有一个明显的优势,即奖励不会像惩罚那样带来无谓的损失。然而,公共品博弈中奖励和惩罚制度在福利方面有效性的比较还是存在争议。大量的文献认为在惩罚制度下平均贡献水平较高,这是因为只有惩罚可信且真实发生,合作才能得以维持。然而,由于惩罚的成本是一个无谓损失,惩罚下的平均收益不会高于奖励下的平



均收益。进一步,在维持合作时,惩罚的优势在不同环境下可能既不显著也不稳健。例如,与大多数实验设置不同,Rand et al. (2009)<sup>[9]</sup>给被试者分配固定的身份,并允许目标互动和直接对等性,这将导致奖励如惩罚一样提升贡献。因此,奖励和惩罚作为一种激励如何起作用 and 如何改变人们的信念和行为,依然存在许多需要阐明的问题,基于更加复杂的带有社会形象考量的社会偏好理论和实验研究可能会提供一些有洞察力的观点。

本文也探讨了公共品博弈中从外生惩罚制度到内生惩罚制度变化的趋势,所出现的考虑内生制度安排的文献聚焦于公共品供给的民主式参与,但这仅是一个开始,更多的研究将会出现在实验室和田野实验中。田野实验在近十年发展迅速,将会是实验室控制实验的有效补充。

#### 【注】

①若没有特殊规定,本文 VCM 所描述的公共品供给机制中,参与者在没有任何激励的情况下根据自己的意愿来供给公共品。

②自愿供给机制下参与者对于公共品供给的动力显而易见的与利他、对等性、关心社会福利等社会偏好有关(Charness 和 Rabin, 2002)。

③在公共品的语境下,“挤出”效应强度与生产中惩罚者的地位相关联。如果惩罚者为一个中央计划者,或者为团队的雇主,那么挤出的强度就更剧烈。相反,如果惩罚者作为社群内的一份子,挤出的强度就没有那么剧烈。

④例如,在真实生活中,我们不能仅仅对别人好,我们也要让被人知道自己对别人好。

#### 参考文献:

- [1] Fehr Ernst, Simon Gächter. Cooperation and Punishment in Public Goods Experiments[J]. American Economic Review, 2000, 90(4): 980 - 994.
- [2] Casari, Marco, Luigi Luini. Peer Punishment in Teams: Expressive Or Instrumental Choice? [J]. Experimental Economics, 2012, 15(2): 241 - 259.
- [3] Nikiforakis, Nikos. Punishment and Counter - Punishment in Public Good Games: Can we Really Govern Ourselves? [J]. Journal of Public Economics, 2008, 92(1 - 2): 91 - 112.
- [4] Herrmann, Benedikt, Christian Thöni, Simon Gächter. Antisocial Punishment Across Societies[J]. Science, 2008, 319

(5868): 1362 - 1367.

[5] Bowles, Samuel, Sung - Ha Hwang. Social Preferences and Public Economics: Mechanism Design When Social Preferences Depend On Incentives[J]. Journal of Public Economics, 2008, 92(8 - 9): 1811 - 1820.

[6] Masclet, David, Charles Noussair, Steven Tucker, Marie - Claire Villeval. Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism[J]. The American Economic Review, 2003, 93(1): 366 - 380.

[7] Andreoni, James, William Harbaugh, Lise Vesterlund. The Carrot Or the Stick: Rewards, Punishments, and Cooperation[J]. American Economic Review, 2003, 93(3): 893 - 902.

[8] Sefton, Martin, Robert Shupp, James M. Walker. The Effect of Rewards and Sanctions in Provision of Public Goods [J]. Economic Inquiry, 2007, 45(4): 671 - 690.

[9] Rand, David G., Anna Dreber, Tore Ellingsen, Drew Fudenberg, and Martin A. Nowak. Positive Interactions Promote Public Cooperation [J]. Science, 2009, 325 (5945): 1272 - 1275.

[10] Bardhan, Pranab. Irrigation and Cooperation: An Empirical Analysis of 48 Irrigation Communities in South India [J]. Economic Development and Cultural Change, 2000, 48 (4): 847 - 865.

[11] Dal B, Pedro, Andrew Foster, Louis Putterman. Institutions and Behavior: Experimental Evidence On the Effects of Democracy[J]. American Economic Review, 2010, 100(5): 2205 - 2229.

[12] Kosfeld, Michael, Akira Okada, Arno Riedl. Institution Formation in Public Goods Games[J]. The American Economic Review, 2009, 99(4): 1335 - 1355.

[13] Walker, James M., Roy Gardner, Andrew Herr, and Elinor Ostrom. Collective Choice in the Commons: Experimental Results on Proposed Allocation Rules and Votes[J]. The Economic Journal, 2000, 110(460): 212 - 234.

[14] Kroll, Stephan, Todd L. Cherry, Jason F. Shogren. Voting, Punishment, and Public Goods[J]. Economic Inquiry, 2007, 45(3): 557 - 570.

[15] Ertan, Arhan, Talbot Page, Louis Putterman. Who to Punish? Individual Decisions and Majority Rule in Mitigating the Free Rider Problem[J]. European Economic Review, 2009, 53 (5): 495 - 511.

[16] Sutter, Matthias, Stefan Haigner, Martin G. Kocher.



Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations[J]. The Review of Economic Studies, 2010, 77(4):1540 – 1566.

[17] Gülerk, özgül, Bernd Irlenbusch, Bettina Rockenbach. Motivating Teammates: The Leader's Choice Between Positive and Negative Incentives[J]. Journal of Economic Psychology, 2009, 30(4): 591 – 607.

[18] Noussair, Charles, Steven Tucker. Combining Monetary and Social Sanctions to Promote Cooperation[J]. Economic Inquiry, 2005, 43(3):649 – 660.

[19] Hopfensitz, Astrid, Ernesto Reuben. The Importance of Emotions for the Effectiveness of Social Punishment[J]. The Economic Journal, 2009, 119(540):1534 – 1559.

[20] Fischbacher, Urs, Simon Gächter. Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments[J]. American Economic Review, 2010, 100(1):541 – 556.

[21] Page, Talbot, Louis Putterman, Bulent Unel. Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry and Efficiency[J]. The Economic Journal, 2005, 115

(506):1032 – 1053.

[22] McEvoy, David M., Todd L. Cherry, John K. Stranlund. The Endogenous Formation of Coalitions to Provide Public Goods: Theory and Experimental Evidence[J]. University of Massachusetts Amherst, 2011, Working Paper Series, No. 2011 – 2.

[23] Charness, Gary, Yang Chun – Lei. Endogenous Group Formation and Public Goods Provision: Exclusion, Exit, Mergers, and Redemption[J]. University of California at Santa Barbara Working Papers Series, 2008.

[24] Maier – Rigaud, Frank P., Peter Martinsson, Gianandrea Staffiero. Ostracism and the Provision of a Public Good: Experimental Evidence[J]. Journal of Economic Behavior & Organization, 2010, 73(3):387 – 395.

[25] Gülerk, özgül, Bernd Irlenbusch, Bettina Rockenbach. The Competitive Advantage of Sanctioning Institutions[J]. Science, 2006, 312(5770):108 – 111.

(责任编辑:宋 敏)

## How Institutions Improve Social Cooperation

### ——A Review of Frontier Theory and Experimental Evidence of Public Goods Game

ZHANG Ping<sup>1</sup>, XU Zhicheng<sup>2</sup>, YAN Jia<sup>1</sup>

(1. China Center for SEZs Research, Shenzhen University, Shenzhen 518060, China;

2. Department of Economics, Texas A&M University, Texas 77843, USA)

**Abstract:** Public goods game, as a fundamental model that studies the coordination and cooperation in human society, has become the focus and difficulty in economics, management, political science, anthropology and other interdisciplinary researches. This paper, based on an analysis of the experimental study on public goods game introduced with incentive system, compares the respective effectiveness of reward and punishment on both individual belief and cooperative supply of public goods game. The existing experimental evidence of public goods game is summarized as external motivations and internal motivations (such as the system of the rules, strength and implement of reward and punishment), and monetary incentives and non – monetary incentives (such as moral persuasion, moral rebukes and social exclusion). The paper explores the actual effect of different institutions on enhancing the cooperative level of public goods supply, with the social net efficiency of the punishment system particularly studied. Based on the conclusions and general implications of the existing literature, the paper proposes possible innovational topics on the theoretical and empirical researches on individual preferences, social co – operations and institutional development.

**Key Words:** public goods; cooperation; reward; punishment; endogenous institution