

惩罚和社会价值取向对公共物品两难中人际信任与合作行为的影响^{*}

王 沛 陈 莉

(上海师范大学教育学院心理学系, 上海 200234)

摘 要 通过两个计算机情境模拟实验, 采用“取消惩罚”范式, 引入社会价值取向变量, 发现惩罚对人际信任和合作行为具有消极影响, 具体表现为当惩罚取消后, 经历过惩罚的被试的人际信任水平显著低于无惩罚条件被试的水平。惩罚对博弈者合作行为的影响在社会价值取向不同的博弈者之间存在显著差异: 经历过惩罚的亲社会型被试在惩罚取消阶段的合作程度显著低于惩罚存在阶段的合作程度, 并且显著低于无惩罚条件被试的相应水平。惩罚通过亲社会型博弈者的人际信任水平对合作程度产生间接负效应, 即惩罚程度越强, 亲社会型博弈者的人际信任水平越低, 进而使其合作程度也下降。

关键词 公共物品两难; 惩罚; 人际信任; 社会价值取向; 合作行为

分类号 B849: C91

1 问题提出

随着科技的进步和竞争的加剧, 个体能力的局限性日渐突出, 几乎每项重要任务的完成, 都离不开合作。然而, 集体中的个体常常要面对集体利益与个体利益相矛盾的局面。例如, 完成团队任务时, 个体在尽己之力和坐享其成之间往往会犹豫不定; 填写个人所得税税单时, 在如实上报和隐瞒真实收入之间常常会徘徊挣扎……Dawes (1980)将诸如此类的两难选择称作社会两难。对于团队中处于社会两难的个体而言, “搭便车”享受他人的贡献比选择合作更有利可图。然而, 这种获益会受到国际合作程度的制约。国际合作程度越高, 个体获益越大 (Messick & Brewer, 1983; Kopelman, Weber, & Messick, 2002)。就其基本属性而言, 社会两难表现出以下特点: 无论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; 行为决策的结果依赖于他人的行为选择; 与合作行为相比, 不合作总是对他人利益有害; 组织内所有成员合作要比他(她)们都不合作使得每个成员获益更多

(Liebrand, Messick, & Wilke, 1992; 刘长江, 李岩梅, 李纾, 2007)。

社会两难的属性表明, 社会两难中的人际合作程度越高, 组织获益越大, 且组织中的个体获益也越大。当前在心理学和经济学的实验研究中, 研究者通常设计一定的游戏或计算机模拟情境来考察社会两难中合作行为的机制, 其中主要的实验范式是公共资源两难和公共物品两难。公共物品两难(public-goods dilemmas)在实验心理学中常常以“给-游戏”(give-some games)的形式进行情境模拟。在该两难范式中, 人们需要进行捐献, 但不要求所有人必须捐献。当捐献达到一定量后便会产生一种公共物品, 并且所有人都可以免费获取该资源。此时, 公共物品的非排除性会导致搭便车现象, 即不创造或维持这种公共物品也可以享用它。在人力资源管理中, 这种现象较为常见。如何减少组织中的“搭便车”行为, 成为组织行为学关注的焦点。

早期研究表明, 惩罚会加强社会两难中的合作 (Caldwell, 1976; Ostrom, Walker, & Gardner, 1992;

收稿日期: 2009-04-22

^{*} 上海市重点学科建设项目(编号: S30401)。

通讯作者: 王沛, E-mail: wangpei1970@163.com

McCusker & Carnevale, 1995; Van Vugt & De Cremer, 1999), 因为惩罚改变了任务结构中局中人的支付结构(Wit & Wilke, 1990; Yamagishi, 1986a, 1986b, 1992; 黄湛冰, 万迪昉, 2005)。Yamagishi (1992)的研究表明: 当群体中的成员相互之间的合作预期较少时, 人们倾向于设置或者引入某种惩罚制度(sanctioning system)。沿着这条线索, 不少研究者提出, 惩罚能有效地增加社会两难中对他人合作的预期。Hobbes 认为, 惩罚制度增加合作的原因有两个: (1)惩罚制度是一种强制力量。它对背信的制裁, 使得背信行为成为不具吸引力的选择; (2)惩罚制度对背信行为的制裁增加了人们对他人合作的预期。近来, 更多的研究表明惩罚是一把“双刃剑”。尽管惩罚可直接促成社会两难中的合作, 但是惩罚对社会两难中决策者的认知框架有着消极影响。具体而言, 惩罚可以使人们把驱使行为的道德动机转变为赢利动机, 致使符合道义的行为极不稳定(Gneezy & Rustichini, 2000; Fehr & Falk, 2001; Mulder, Van Dijk, De Cremer, & Wilke, 2005)。然而, 在社会两难中, 成员之间是相互依赖的关系。个体的合作行为不仅仅受到自己合作动机的影响, 还依赖于对他人合作的预期。若一定的情境线索使得博弈者预期其他人会合作, 则博弈者会表现出更多的合作意愿和合作行为; 若一定的情境线索使得博弈者不相信其他人会合作, 则博弈者会因避免他人剥削, 进而采取竞争行为。国外已有研究发现, 个体对他人合作行为的预期是影响合作的重要中介变量之一(Yamagishi, 1986a, 1986b; Robbins, 1995; De Cremer, Dewitte, & Snyder, 2001; De Cremer & Van Dijk, 2002)。国内除有学者从演化博弈的角度提出, 当人际信任程度较高时, 增加惩罚力度才会增强合作, 但是鲜有实证研究的证据(张朝孝, 2006)。

信任有许多不同的定义, 但总体上可以归为两大类: 突出社会道德层面的信任(人际信任)和强调利益算计层面的信任(条件信任)(Blomqvist, 1997)。在已有的社会两难研究中, 信任是指博弈者对他人合作行为的预期。Sitkin 和 Roth (1993)指出, 人们可能会因为惩罚等强制措施的出现而形成信任, 并把这种信任称为“制度信任”(institutional trust)。根据信任的区分定义, 制度信任实质上是一种条件信任。Sitkin 和 Roth (1993)发现惩罚可以激发条件信任。然而, 来自社会两难情境以外的研究表明, 惩罚会破坏人际信任(De Dreu, Giebels, &

Van de Vliert, 1998; Fehr & Gächter, 2002)。人际信任(interpersonal trust)是基于对对方意图和行为的积极预期, 愿意向对方暴露自己的弱点并且不担心被利用的一种心理状态。人际信任包含两个要素: 信任意向(trusting intention, 愿意依靠对方)和信任信念(trusting belief, 相信对方是仁慈的、诚实的、有能力的和可预测的)(Mayer, Davis, & Schoorman, 1995; Rousseau, Sitkin, & Burt, 1998)。Mulder 等人(2006)使用“取消惩罚”的实验范式研究发现: 通过削弱博弈者的人际信任水平, 惩罚减弱了博弈者的合作程度。研究还表明, 博弈者基于惩罚产生条件信任从而加强合作, 然而条件信任会随着惩罚的取消而消失。无论惩罚破坏合作, 还是增加合作, 惩罚均会影响合作, 并且分别表现为两个效应: 一个是信号效应(signaling effect), 此时惩罚可以影响人们决策结构的类型。与伦理决策结构(ethical decision frame)相比较, 惩罚的存在增加了交易结构(business frame)被唤起的可能性; 另一个叫处理效应(processing effect), 此时决策结构影响着人们策略的选择, 不同的决策结构导致不同的处理价值(processing value)。在伦理决策结构中, 不管惩罚的大小, 合作是主要的选择。但是在交易结构中, 成本效应分析过程则会考虑到惩罚的大小, 合作是利益分析后的选择(陈欣, 叶浩生, 2007)。

综上所述, 惩罚会影响合作和人际信任, 人际信任也会影响合作水平。然而, 除 Mulder 等人(2006)发现惩罚对合作的影响可以通过人际信任这一中介变量起作用以外, 对惩罚、人际信任和合作三者之间的关系还鲜有研究。Yamagishi (1988a, 1988b)的研究发现: 在没有惩罚的社会两难中, 日本被试的合作水平显著低于美国被试。李纾(2005, 2006)的研究表明: 个人主义文化和群体主义文化中群体的合作决策必然会表现出一定的差异, 有可能惩罚通过影响人际信任水平而影响合作的效应会因为被试社会价值取向的不同而不同。因此, 本研究采用 Mulder 等人(2006)创设的“取消惩罚”范式(RTS, Removing The Sanction), 选用中国大学生为被试, 通过计算机模拟两难情景, 首先对 Mulder 等人的研究结论进行验证。在此基础上, 引入社会价值取向这一变量, 考察惩罚在社会两难中对博弈者的人际信任水平和合作水平的影响, 以期进一步探明惩罚影响合作行为的心理机制, 并为人力资源管理者提出合理利用惩罚的新途径。

2 实验 1 惩罚对公共物品两难中博弈者人际信任和合作水平的影响

2.1 目的

通过操纵实验条件,考察公共物品两难中惩罚对博弈者人际信任和合作行为的影响,同时验证 Mulder (2006)等人通过设计消除惩罚的范式(RTS, Removing The Sanction)研究所得的结论。研究假设为:经历过惩罚的被试总体信任水平和合作水平都显著下降。

2.2 方法

2.2.1 被试 招募自愿参加实验的大学生 76 名,男女各半,年龄在 20~30 岁之间。最后收集有效数据 72 人,其中男性 32 人。

2.2.2 实验设计 采用 2(惩罚条件:无惩罚,有惩罚) $\times$ 2(实验阶段:阶段 1,阶段 2)的混合设计。其中惩罚条件为组间变量,因变量为被试在局域网的虚拟情境中实际贡献的网时和对他人合作行为的信任程度。对他人合作行为的信任程度由题目“你相信其他小组成员会为组织利益贡献个人网时吗?(1 完全不相信,2 不相信,3 有点不信,4 一般,5 有点相信,6 相信,7 完全相信)”进行测量(Mulder et al, 2006)。所测到的信任为 Mulder 等人(2006)的研究所主张的总体信任水平,即指博弈者对他人合作行为的预期。

2.2.3 实验程序 采用 Visual basic 6.0 编程。一次同时安排 4 名被试为一个小组完成实验,每名被试在独立小隔间内完成整个实验。指导语为:“您将与同时被安排进入实验室的其他三名成员形成一个小组。在完成以下两个阶段任务时,相互之间不能交流。您可能在第一阶段获得额外奖励,也可能在第二阶段获得额外奖励。”

第一阶段:1、情境:小组中的每个成员都将免费获得一张 30 小时的校园网网卡。当将网卡上免费赠送的网时用尽时,您需要自费充值。但您可以转移自己的网时到公共电子充值账户中为小组共有。公共电子充值账户接受转账后即刻升值,将会使小组中的公共网时增加一倍。增值后的公共网时将会平均分配到每个小组成员的个人网卡充值账户里。分配在有惩罚条件下的被试被告知:对于小组中贡献网时最少的成员将扣除其网卡上所有的网时。提示:如果小组中的每个成员都自愿贡献网时,则相应的每个成员网卡上的网时都会增加。

如果只有你一人贡献,其他三人不贡献,你贡献一小时网时虽然增值一倍变为两小时,但被小组中的四个成员平分后,你将因此而损失 0.5 小时网时。2、任务:(1)你相信其他小组成员会为组织利益贡献个人网时吗?(1 完全不相信,2 不相信,3 有点不信,4 一般,5 有点相信,6 相信,7 完全相信);(2)你决定贡献多少网时给你所在的小组?(用 0~30 以内的具体数字表示)。

第二阶段情境与第一阶段相同,只是分配在惩罚条件下的被试被告知惩罚取消,且阶段 2 的决策与他(她)们在阶段 1 的决策无关。为了使被试评价实验阶段 1 和阶段 2 为相互独立的情景,告知被试获得的网时可能会在阶段 1 也可能在阶段 2 转移到小组公共电子充值账户上。

在被试完成第二阶段的任務后,对每名参与实验的被试给予奖励,并兑现承诺,即随机选择被试给予额外的奖励。实验结束后,就实验和被试进行一个简短的访谈,如果发现有的被试对实验情境产生疑虑,数据作废。最后对被试解释整个实验过程,并请被试为本实验保密。

2.3 结果与分析

2.3.1 惩罚条件对博弈者在公共物品两难情景中信任的影响 为分析惩罚条件对博弈者在公共物品两难情景中信任的影响,在两个实验阶段,分别对处于不同惩罚条件下的被试对他人合作的信任程度的平均值和标准差进行描述统计,结果见表 1。

表 1 不同惩罚条件的博弈者在实验两个阶段对他人合作的信任程度

惩罚条件	阶段 1			阶段 2		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
无惩罚	4.28	1.54	36	4.11	1.56	36
有惩罚	5.47	0.97	36	3.25	1.05	36

以惩罚条件和实验阶段为自变量对信任度进行 2 $\times$ 2 的重复测量方差分析。结果发现:惩罚条件的主效应不显著, $F(1,70)=0.32$, $p=0.58$, 效应量 $\eta^2=0.01$; 实验阶段的主效应显著, $F(1,70)=183.33$, $p<0.001$, 效应量 $\eta^2=0.72$; 更重要的是,惩罚条件与实验阶段的交互作用显著, $F(1,70)=135.74$, $p<0.001$, 效应量 $\eta^2=0.07$ 。分别在惩罚条件的两个水平上对实验阶段进行简单效应检验,结果表明:在有惩罚条件下,实验阶段效应显著, $F(1,70)=317.28$, $p<0.001$, 效应量 $\eta^2=0.82$, 即同阶段 1 的被试相比,阶段 2 的被试的信任水平更低。然而,当无惩罚条

件时, 实验阶段效应不显著, $F(1,70)=1.18, p=0.19$, 效应量 $\eta^2=0.03$ 。在实验阶段的两个水平上对惩罚条件进行简单效应检验, 结果发现: 在阶段 1, 惩罚条件的效应显著, $F(1,70)=1030.85, p<0.001$ 。同无惩罚条件的被试相比, 惩罚条件下的被试的信任水平更高。然而, 在实验阶段 2, 惩罚条件下的被试与无惩罚条件下的被试的信任水平差异显著, $F(1,70)=549.25, p<0.001$ 。

2.3.2 惩罚对博弈者在公共物品两难情景中合作行为的影响 为检验惩罚对博弈者合作行为的影响, 以惩罚条件和实验阶段为自变量对合作程度进行 2×2 的重复测量方差分析(描述统计见表 2)。结果表明: 实验阶段的主效应不显著, $F(1,70)=2.94, p=0.091$, 效应量 $\eta^2=0.04$ 。而实验阶段与惩罚条件的交互作用显著, $F(1,70)=19.34, p<0.001$, 效应量 $\eta^2=0.36$ 。

表 2 不同惩罚条件的被试在实验两阶段的合作程度

惩罚条件	阶段 1			阶段 2		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
无惩罚	10.67	6.80	36	12.47	6.84	36
有惩罚	15.94	8.19	36	11.83	6.51	36

对实验阶段和惩罚条件的交互作用进一步做简单效应分析, 结果表明, 有惩罚条件下, 实验阶段的效应显著, $F(1,70)=18.67, p<0.001$, 同阶段 1 的被试相比, 在实验阶段 2, 被试的合作程度更低。然而, 在无惩罚条件时, 实验阶段的效应不显著, $F(1,70)=3.60, p=0.062$ 。由此可见, 在阶段 2, 仅仅是经历过惩罚的博弈者的合作程度较之阶段 1 有所下降。

为了考察惩罚取消阶段博弈者的合作程度是否低于无惩罚条件组被试的水平, 分别在实验阶段的两个水平上对惩罚条件进行简单效应检验, 结果表明: 在阶段 2, 惩罚条件效应不显著, $F<1$, 即分配在有惩罚条件组被试的合作程度 ($M=11.83, SD=6.51$) 并不显著低于无惩罚条件组被试的合作程度 ($M=12.47, SD=6.84$)。

2.3.3 博弈者的信任水平与合作水平的相关分析 研究一的结果表明, 在阶段 2, 惩罚对博弈者的信任水平有显著影响, 但对其合作水平并无显著影响。这可能意味着在阶段 2 博弈者的合作程度与其信任水平并不相关。然而, 对阶段 2 博弈者的信任水平和合作程度进行相关分析发现: 在阶段 2, 博弈者的信任水平与合作程度显著相关 ($r=0.33$,

$p=0.005$)。

2.4 讨论

本实验的结果表明: 当已经存在的惩罚取消后, 博弈者的信任水平和合作程度均有所下降。然而在实验第二阶段, 随着惩罚的取消, 博弈者基于惩罚产生的信任已经消失, 此时的信任应该主要是人际信任。既然在实验第二阶段, 经历过惩罚条件的博弈者的信任水平显著低于无惩罚条件被试的信任水平, 由这一结果可以推断惩罚削弱了博弈者的人际信任水平。以上结果与研究预期一致, 即惩罚对处于公共物品两难中的博弈者的人际信任水平将会产生消极的影响。博弈者可能相信只要惩罚制度存在, 小组中的其他成员会因避免惩罚而采取合作。惩罚取消后, 经历过惩罚且产生条件信任的被试的信任水平则会低于在阶段 1 中没有经历过惩罚的被试。这些研究推断与 Mulder 等人(2006)的研究结论想吻合。即在实验阶段 1 中, 惩罚的存在破坏了博弈者的人际信任水平。

然而, 尽管在阶段 2, 博弈者的合作程度也有所下降, 但在实验第二阶段, 经历过惩罚条件的被试的合作程度并未显著低于未经历过惩罚条件的被试。这一结果表明, 博弈者的人际信任水平与其合作行为可能无关, 或受到了其他因素的影响。为此很有必要进一步探讨博弈者的人际信任与其合作行为之间的关系。

3 实验 2 惩罚对公共物品两难中不同社会价值取向的博弈者的人际信任水平和合作水平的影响

3.1 目的

Yamagishi (1986a, 1986b) 发现, 惩罚与人际信任的关系可能还依赖于人们最初的人际信任倾向。在某种社会两难中, 如果一开始人们对其他人会合作持怀疑态度, 惩罚体制的设置就是必要的, 即惩罚可能促进低人际信任倾向个体的信任水平。然而, 研究一的结果表明, 惩罚削弱了公共物品两难中博弈者的人际信任水平, 这一结果是否在具有不同社会价值取向的被试中也适用, 有待于在引入社会价值取向这一变量后进一步验证。为此, 实验二进一步探明惩罚、人际信任与合作行为三者之间的关系。

Liebrand 等人(1986)认为个人的社会价值取向有效地预测了个人的人际信任倾向, 亲社会型价值取向的人具有高人际信任倾向, 个体价值取向的人

具有低人际信任倾向。并且提出个人的社会价值取向影响了其对于人际间相互依赖的理解,亲社会型的个体对于混合动机冲突中的竞争合作维度更多的是从道德动机的角度理解,而竞争型的个体则更注重权利、利益等外部动机。Mulder (2006)的研究表明惩罚会破坏人际信任。由此可以推断:随着惩罚的撤销,条件信任随即消失。对于初始人际信任水平低的个体价值取向的被试而言,即使在实验第一阶段因为惩罚产生条件信任,但这种信任在实验第二阶段会随着惩罚的取消而消失,而个体本身的人际信任水平基本不受影响。但是,对于初始人际信任水平高的亲社会型个体而言,由于惩罚的存在可以看作是组织内部不可信任的线索,所以惩罚的取消会使得被试难以相信组织成员之间会自觉合作(Tenbrunsel & Messick, 1999; Fehr & Falk, 2001)。由此实验二的假设为:惩罚条件对不同社会价值取向的博弈者的人际信任的影响有显著差异。在实验阶段 2, 经历不同惩罚条件的亲社会价值取向的被试的人际信任水平有显著差异,而经历不同惩罚条件的个体价值取向的被试的人际信任水平无显著差异。

已有研究表明惩罚影响社会两难中博弈者的人际信任水平(De Dreu, Giebels, & Van de Vliert, 1998; Fehr & Gächter, 2002), 而博弈者在社会两难中的合作程度受到博弈者人际信任水平的影响(De Cremer & Van Dijk, 2002; Mulder, Van Dijk, & Wilke, 2006)。此外, 研究一表明在实验阶段 2, 博弈者的信任水平和合作水平显著相关, 实验二预期惩罚通过影响不同社会价值取向博弈者的信任水平进而影响其合作水平, 即经历不同惩罚条件的亲社会价值取向的被试的合作水平有显著差异, 而经历不同惩罚条件的个体价值取向的被试的合作水平无显著差异。

3.2 方法

3.2.1 被试 以付酬的形式主要通过校园网络招纳自愿报名的大学生被试共 126 名。通过社会价值取向问卷的测量结果确定参加正式实验的被试为 96 名, 亲社会价值取向的被试和个体价值取向的被试均为 48 名。其中参加正式实验的男性 46 人, 女性 50 人, 年龄为 20~30 岁。

3.2.2 实验设计 采用 3(惩罚条件: 无惩罚, 轻度惩罚, 重度惩罚)×2(社会价值取向: 亲社会价值取向, 个体价值取向)×2(实验阶段: 阶段 1, 阶段 2)的混合实验设计, 其中实验阶段为被试内变量, 因

变量为被试的合作程度和人际信任水平。人际信任将通过基于实验程序对总体信任水平的统计比较和推断而来的, 即通过运用 Mulder 等人“取消惩罚”的研究范式, 比较总体信任水平的变化来推断惩罚对人际信任的影响。这种做法和 Mulder 等人(2006)的研究中关于人际信任的测量方法一致。

3.2.3 实验前测 (1) 确定惩罚程度 随机抽取大学生 90 名(男女各半), 参与由计算机呈现的模拟公共物品两难游戏。五种惩罚条件(无惩罚、扣除个人网卡上的 5 小时网时、10 小时网时、15 小时网时、20 小时网时)在计算机屏幕上随机呈现给被试, 为了保证被试在每种惩罚条件下作答的稳定性, 每种惩罚条件随机呈现两次。前测结果如表 3 所示。

表 3 惩罚程度对被试网时的贡献量

惩罚程度	<i>n</i>	<i>M</i>	<i>SD</i>
无惩罚	18	6.33	3.07
扣除 5 小时	18	13.83	3.71
扣除 10 小时	18	14.89	4.40
扣除 15 小时	18	23.00	3.50
扣除 20 小时	18	23.83	3.71
总计	90	16.38	7.44

将每个被试同一惩罚程度下两次贡献的网时取平均, 形成 5 组平均值。以惩罚程度为组内变量, 以不同惩罚程度下平均网时贡献量为因变量进行重复测量方差分析, $F(4,85)=68.52, p<0.001$ 。随着惩罚程度的不断增加, 被试的网时贡献量显著增加。同时, 惩罚程度为“扣除 5 小时”时出现了第一个拐点, 5 小时之后渐趋平稳且与 10 小时之间差异不大, 在 15 小时后被试的网时贡献量又渐趋平稳, 在惩罚程度为 20 小时时, 网时贡献量达到第一次高峰。根据前测结果将惩罚条件分为: 轻度惩罚条件, 对于小组中贡献网时最少的成员, 将扣除其网卡上的 5 小时网时; 重度惩罚条件: 对于小组中贡献网时最少的成员, 将扣除其网卡上的 20 小时网时。

(2) 被试社会价值取向的测量

采用“三重对策矩阵”的方法来评价被试的社会价值取向, 根据社会价值取向问卷的测量结果将被试划分为亲社会价值取向和个体价值取向两组。王重鸣、严进(2000, 2002, 2003)等人多次在其研究中都采用这种方法来测量个体的价值取向。具体测验矩阵如表 4 所示。

表 4 社会价值取向测量矩阵示例

结果	例一		例二		例三	
	自己 获益	他人 获益	自己 获益	他人 获益	自己 获益	他人 获益
选择一	50	20	50	15	60	30
选择二	40	0	40	0	50	10
选择三	40	40	40	40	50	50

以上矩阵的规律为: 选择一的自己获益是 3 个选择中最高的, 选择二的自己获益与他人获益的差异是 3 个选择中最大的, 选择三的双方获益之和是 3 个选择中最高的。测试时将三种选择的顺序随机打乱。不论是被试倾向于个人利益、群体利益、还是与对手的得益差异, 都只能倾向于其中的一种选择。如果倾向于选择一, 被试个人的绝对获益最高, 表明被试是个人主义取向; 倾向于选择二, 表明被试倾向于与别人竞争, 是竞争型的价值取向; 倾向于选择三的被试着重关心双方的共同利益, 这样的被试属于亲社会价值取向。基于全部被试社会价值取向测量的结果, 按照 Dreu 与 McCusker 的标准, 将在 12 次决策中至少有 7 次做出一致决策的被试确定为具有较稳定价值取向的个体, 而且根据本实验的研究目的, 将个人型和竞争型的被试作为一类, 统称为个体价值取向型。统计结果表明, 亲社会价值取向的被试有 72 名, 个体价值取向的被试有 48 名, 另外有 6 名被试的反应不稳定, 故其不参与正式实验, 为使得两类被试数目平衡, 各取 48 名被试参加正式实验。

3.2.4 实验程序 采用 Visual basic 6.0 编程。将两种社会价值取向的被试分别随机分配到无惩罚、轻度惩罚和重度惩罚三种条件下, 同时安排 4 名被试为一个小组完成实验, 每个被试在独立小隔间内完成整个实验。实验第一阶段在电脑上分别为随机分配在无惩罚、轻度惩罚条件和重度惩罚条件下的被试呈现模拟两难情景(同实验 1), 在每一阶段博弈者做完决策后, 对其信任水平和合作水平的测量都与实验 1 相同。第二阶段情境与第一阶段相同, 但通过指导语告知分配在有惩罚条件下的被试, 惩罚已经被取消且惩罚的取消与他(她)们在第一阶段的决策无关。

3.3 结果

3.3.1 社会价值取向、惩罚条件和实验阶段对博弈者在公共物品两难情景中人际信任的影响 在阶段 1 和阶段 2, 经历过不同惩罚条件的不同社会价值取向的博弈者信任水平的平均值和标准差见表

5。可以看出, 经历过重度和轻度惩罚的被试其总体信任水平在惩罚取消阶段有所下降, 而未经历惩罚的被试其信任水平在阶段 2 基本没有大的变化。

表 5 不同价值取向的被试在不同惩罚条件下在两阶段的信任水平($n=16$)

惩罚条件	社会价值取向	实验阶段	<i>M</i>	<i>SD</i>
无惩罚	亲社会价值取向	1	5.31	0.60
		2	5.00	0.97
	个体价值取向	1	4.00	1.32
		2	4.00	1.10
轻度惩罚	亲社会价值取向	1	5.25	0.77
		2	4.25	1.00
	个体价值取向	1	3.75	1.00
		2	3.63	1.02
重度惩罚	亲社会价值取向	1	5.81	0.75
		2	2.88	1.54
	个体价值取向	1	4.13	0.89
		2	3.63	0.96

为了考察同阶段 1 相比, 被试在阶段 2 的信任水平是否更低, 以惩罚条件、社会价值取向和实验阶段为自变量对信任水平进行 $3 \times 2 \times 2$ 的重复测量方差分析, 其中实验阶段为被试内因素。结果显示, 实验阶段的主效应显著, $F(1,90)=71.19$, $p<0.001$, 效应量 $\eta^2=0.44$; 惩罚条件、社会价值取向和实验阶段的三重交互作用显著, $F(2,90)=10.89$, $p<0.001$, 效应量 $\eta^2=0.20$ 。为了考察实验阶段效应在什么情况下出现, 将惩罚条件和社会价值取向的水平固定, 进一步进行实验阶段的简单效应检验。结果显示: 无惩罚条件下, 亲社会价值取向的被试, 实验阶段效应不显著, $F(1,90)=1.76$, $p=0.19$, 个体价值取向的被试, 实验阶段效应也不显著, $F<1$; 轻度惩罚条件下亲社会价值取向的被试, 实验阶段效应显著, $F(1,90)=17.97$, $p<0.001$, 个体价值取向的被试, 实验阶段效应不显著, $F(1,90)=0.28$, $p=0.597$; 重度惩罚条件下, 亲社会价值取向的被试, 实验阶段效应显著, $F(1,90)=155.09$, $p<0.001$, 个体价值取向的被试, 实验阶段效应临界显著, $F(1,90)=4.49$, $p=0.04$ 。

3.3.2 惩罚条件和社会价值取向对博弈者在实验阶段 2 的人际信任的影响 以上分析表明, 在阶段 2, 经历过惩罚的被试的人际信任水平有所下降。接下来以实验阶段 1 和实验阶段 2 的总体信任水平之差作为因变量, 以惩罚条件和社会价值取向作为自变量, 对两实验阶段的总体信任水平之差进行方差分析。不同价值取向的被试在不同惩罚条件下的总体

信任水平之差的描述统计见表 6。表 6 表明亲社会价值取向被试的总体信任水平的下降程度要大于个体价值取向的被试。进一步做 3(惩罚条件)×2(社会价值取向)的多因素方差分析, 结果表明惩罚条件的主效应极其显著, $F(2,90) = 23.62, p < 0.001$, 效应量 $\eta^2 = 0.34$; 价值取向的主效应极其显著, $F(1, 90) = 39.36, p < 0.001$, 效应量 $\eta^2 = 0.30$; 惩罚条件与价值取向的交互作用极其显著, $F(2,90) = 10.89, p < 0.001$, 效应量 $\eta^2 = 0.20$ 。

表 6 不同价值取向的被试在不同惩罚条件下的两阶段的总体信任水平之差($n=16$)

社会价值取向	惩罚条件	<i>M</i>	<i>SD</i>
亲社会价值取向	无惩罚	0.31	0.60
	轻度惩罚	1.00	0.82
	重度惩罚	2.94	1.61
个体价值取向	无惩罚	0.00	0.82
	轻度惩罚	0.13	0.62
	重度惩罚	0.50	0.82
总体	亲社会价值取向	1.42	1.56
	个体价值取向	0.21	0.77

在价值取向的不同水平上, 对惩罚条件的简单效应检验结果表明, 惩罚条件对于亲社会价值取向的被试有着非常显著的影响, $F(2,90)=33.30, p < 0.001$, 效应量 $\eta^2=0.43$, 进一步的多重比较表明, 重度惩罚条件下人际信任水平的下降程度($M=2.938$)显著大于无惩罚($M=0.313$)和轻度惩罚条件下($M=1.000$)人际信任的下降程度, 轻度惩罚条件下的人际信任下降程度要大于无惩罚条件的, 但两者不存在统计学上的显著差异; 而对于个体价值取向的被试而言, 惩罚条件对其没有显著影响, $F(2,90)=1.22, p > 0.05$, 效应量 $\eta^2=0.03$ 。即惩罚程度越大, 对亲社会价值取向的博弈者在实验阶段 2 的人际信任水平削弱的程度越大, 而不同的惩罚条件对个体价值取向的博弈者在实验阶段 2 的人际信任水平无显著影响。

3.3.3 社会价值取向、惩罚条件和实验阶段对博弈者在公共物品两难情景中合作水平的影响 在阶段 1 和阶段 2, 经历过不同惩罚条件的不同社会价值取向的博弈者合作水平的平均值和标准差见表 7。由表 7 可以看出, 在实验阶段 2, 经历过重度惩罚的博弈者合作水平下降水平最为明显, 经历轻度惩罚条件者次之, 但也有明显下降; 而没有经历过惩罚条件的被试合作水平基本没有大的变化。

表 7 不同价值取向的被试在不同惩罚条件下的两阶段的合作水平($n=16$)

惩罚条件	社会价值取向	实验阶段	<i>M</i>	<i>SD</i>
无惩罚	亲社会价值取向	1	14.63	6.96
		2	14.31	6.09
轻度惩罚	亲社会价值取向	1	9.44	4.57
		2	9.25	4.71
	个体价值取向	1	10.75	4.75
		2	8.69	2.15
重度惩罚	亲社会价值取向	1	24.25	4.70
		2	8.06	2.86
	个体价值取向	1	14.38	3.40
		2	11.50	3.25

为了考察被试在阶段 2 的合作水平是否比阶段 1 更低, 以惩罚条件、社会价值取向和实验阶段为自变量对合作水平进行 $3 \times 2 \times 2$ 的重复测量方差分析, 其中实验阶段为被试内因素。结果显示, 实验阶段的主效应显著, $F(1,90)=159.70, p < 0.001$, 效应量 $\eta^2=0.64$; 惩罚条件、社会价值取向和实验阶段的三重交互作用显著, $F(2,90)=25.29, p < 0.001$, 效应量 $\eta^2=0.36$ 。为了考察实验阶段效应在什么情况下出现, 将惩罚条件和社会价值取向的水平固定, 进一步进行实验阶段的简单效应检验。结果显示: 无惩罚条件下, 亲社会价值取向的被试, 实验阶段效应不显著, $F(1,90)=0.11, p=0.74$, 个体价值取向的被试, 实验阶段效应也不显著, $F(1,90)=0.04, p=0.84$; 轻度惩罚条件下亲社会价值取向的被试, 实验阶段效应显著, $F(1,90)=61.19, p < 0.001$, 个体价值取向的被试, 实验阶段效应不显著, $F(1,90)=4.87, p=0.03$; 重度惩罚条件下, 亲社会价值取向的被试, 实验阶段效应显著, $F(1,90)=299.84, p < 0.001$, 个体价值取向的被试, 实验阶段效应也显著, $F(1,90)=9.46, p=0.003$ 。

3.3.4 惩罚条件和社会价值取向在实验阶段 2 对博弈者合作水平的影响 以上分析表明, 在实验阶段 2, 经历过惩罚的被试的合作水平有所下降, 以实验阶段 1 和实验阶段 2 的合作水平之差即合作水平的下降程度为因变量, 以惩罚条件和社会价值取向为自变量对两阶段的合作水平之差进行方差分析。不同价值取向的被试在不同惩罚条件下的合作水平的下降程度的描述统计见表 8。表 8 表明亲社会价值取向被试的合作水平的下降程度要大于个体

价值取向的被试。进一步做 3(惩罚条件) $\times$ 2(社会价值取向)方差分析, 结果表明惩罚条件的主效应极其显著, $F(2,90)=49.32$, $p<0.001$, 效应量 $\eta^2=0.52$; 价值取向的主效应极其显著, $F(1,90)=66.60$, $p<0.001$, 效应量 $\eta^2=0.43$; 惩罚条件与价值取向的交互作用极其显著, $F(2,90)=25.29$, $p<0.001$, 效应量 $\eta^2=0.36$ 。

表 8 不同价值取向的被试在不同惩罚条件的两实验阶段的合作水平的下降程度($n=16$)

社会价值取向	惩罚条件	M	SD
亲社会价值取向	无惩罚	0.31	4.64
	轻度惩罚	7.31	4.32
	重度惩罚	16.19	4.22
个体价值取向	无惩罚	0.19	1.05
	轻度惩罚	2.06	3.38
	重度惩罚	2.88	3.67
总体	亲社会价值取向	7.94	7.85
	个体价值取向	1.71	3.09

在价值取向的不同水平上, 对惩罚条件的简单效应检验结果表明, 惩罚条件对于亲社会价值取向的被试有非常显著的影响, $F(2,90)=72.43$, $p<0.001$, 效应量 $\eta^2=0.62$, 进一步的多重比较表明, 重度惩罚条件下人际信任水平的下降程度($M=2.88$)显著大于轻度惩罚($M=2.06$)和无惩罚($M=0.19$)条件的人际信任水平的下降程度, 轻度惩罚条件的人际信任下降程度也显著大于无惩罚条件的; 而对于个体价值取向的被试而言, 惩罚条件对其合作行为没有显著影响, $F(2,90)=2.17$, $p>0.05$, 效应量 $\eta^2=0.05$ 。也就是说, 惩罚的程度越大, 对亲社会型博弈者在惩罚取消阶段的合作行为的削弱程度越大, 而不同的惩罚条件对个体价值取向的博弈者在惩罚取消阶段的合作行为无显著影响。这些研究结果证实了研究最初的假设, 即惩罚的存在对亲社会型价值取向博弈者的合作行为有消极影响。

3.3.5 人际信任在惩罚影响博弈者合作行为中的中介作用 按照中介效应检验的程序(温忠麟, 张雷, 侯杰泰, 刘红云, 2004)对人际信任(P)的中介效应依次检验, 具体程序如下:

(1) 在实验阶段 1, 惩罚的存在对实验阶段 2 中个体价值取向型博弈者合作行为的预测以及人际信任在惩罚影响其合作行为中的中介效应检验。对于个体价值取向的博弈者, 以惩罚为预测因素对

其合作程度进行回归分析。结果表明惩罚没有显著预测博弈者的合作程度 $\beta=0.25$, $p=0.082$ 。因此, 对于个体价值取向的博弈者已经可以推断出不存在人际信任的中介效应, 因为惩罚条件并没有影响第二阶段博弈者的合作程度, 因此没有必要再进一步检测人际信任的中介效应。

(2) 在实验阶段 1, 惩罚的存在对实验阶段 2 中亲社会型博弈者合作行为的预测以及人际信任在惩罚影响亲社会型博弈者合作行为中的中介效应检验。研究转向于检验人际信任在惩罚对亲社会价值取向博弈者合作行为影响中的中介效应。首先, 将惩罚作为预测因素对亲社会型博弈者的合作程度进行回归分析。研究表明惩罚在阶段 1 的存在预测了实验阶段 2 中博弈者的合作行为 $\beta = -0.44$, $p = 0.002$ 。

以惩罚作为预测因素, 对实验第二阶段亲社会型博弈者的人际信任度进行回归分析表明实验第一阶段存在的惩罚预测了实验第二阶段的人际信任, $\beta = -0.60$, $p<0.001$ 。

人际信任(P)的中介效应分析结果见表 9, 其中的结果是标准化解, 用小写字母代表相应变量的标准化变量。由于依次检验(前面 3 个检验)都是显著的, 所以人际信任的中介效应显著。由于第四个 t 检验不显著, 所以是完全中介效应, 中介效应占总效应的比例为 $0.595 \times 0.588 / 0.444 = 78.8\%$ 。

表 9 人际信任(P)的中介效应依次检验

	标准化回归方程	回归系数检验
第一步	$y = -0.444x$	$SE = 0.931$, $t = -3.357^{**}$
第二步	$P = -0.595x$	$SE = 0.211$, $t = -5.026^{***}$
第三步	$y = 0.588p$	$SE = 0.828$, $t = 2.207^*$
	$-0.094x$	$SE = 1.419$, $t = -.533$

注: SE 表示标准误, $^*p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$

上述包含了中介变量 P 的模型分析结果表明: 惩罚通过亲社会型博弈者的人际信任水平对合作程度有间接负效应, 即惩罚程度越强, 在实验阶段 2, 亲社会型博弈者的人际信任水平越低, 而此时亲社会型博弈者的人际信任水平会影响博弈者的合作程度, 使其合作程度也下降。

3.4 讨论

实验二的结果更精确地证实了实验一的结论: 当已经存在的惩罚取消后, 亲社会型博弈者和个体价值取向的博弈者其人际信任程度和合作程度均

有所下降。亲社会型博弈者是具有高人际信任倾向的个体,他(她)们倾向于信任他人的合作行为是基于道德层面的。取消惩罚后,经历过惩罚的博弈者的信任水平显著低于无惩罚条件组被试的水平,这一结果再次证实了研究假设,即惩罚削弱了博弈者的人际信任水平。中介效应分析的结果表明,惩罚通过削弱亲社会价值取向博弈者的人际信任水平导致其合作程度的下降。

此外,实验二的结果表明,经历了重度惩罚条件的亲社会型博弈者在实验阶段2的信任水平和合作水平显著低于轻度惩罚条件和无惩罚条件。这一结果可以从逆推的角度说明惩罚使得博弈者产生条件信任,即惩罚作为一种线索使人们意识到小组中的其他成员不会自觉合作,而是基于惩罚的存在才迫于合作。所以惩罚一旦取消,基于惩罚建立的条件信任都随之消失,以至于经历过重度惩罚条件的被试在惩罚取消后,人际信任水平最低。

Mulder 等人(2006)的研究表明在社会两难中,惩罚通过削弱人际信任水平而削弱合作。但实验二在引入了社会价值取向这一变量对被试进行亲社会价值取向和个体价值取向的区分后,结果发现人际信任在惩罚影响合作中的中介作用仅限于亲社会价值取向的博弈者。这种结果的差异可以在社会两难领域以外的研究中找到证据,按照 Pillutla 和 Chen (1999)提出的观点,可以认为个体取向的被试本身的人际信任倾向就低,而惩罚作为他们不信任的认知线索,使得惩罚提供的信息与他们的预期相一致,所以惩罚对个体取向的被试的信任水平和合作程度的影响程度会低于亲社会型被试。因此,在实验二中引入社会价值取向变量考查人际信任水平对合作程度的影响是有必要的。

4 总讨论

4.1 惩罚与合作

在社会两难领域,许多早期的研究表明惩罚对合作有促进作用,并且相对于重度惩罚,轻度惩罚会导致人们的不合作行为(Caldwell, 1976; Yamagishi, 1986a, 1986b, 1992; McCusker & Carnevale, 1995; Van Vugt & De Cremer, 1999; Eek et al., 2002)。但 Tenbrunsel 和 Messick (1999)的研究表明惩罚对于社会两难情境中决策者的认知决策框架有消极影响。具体研究表明,惩罚程度无论轻重,都会对人们的认知决策框架产生消极影响。但是,本研究实

验二的结果表明,重度惩罚会削弱合作程度。而且本研究通过“取消惩罚”的范式才使这种消极影响在惩罚取消阶段得以呈现。在运用其它范式的研究中也可以找到这种消极影响的证据:当一种惩罚制度出现时,行为被不必要地监控,对他人的不信任可能会驱使被试在没有惩罚的时候采取背叛,特别在有背叛的时机出现时。人们可能寻求可能的背叛方法去逃避惩罚(Mulder, Van Dijk, De Cremer, & Wilke, 2005)。

4.2 惩罚与人际信任

本研究表明仅仅是一种惩罚制度的出现就会使人们依赖惩罚制度:因为不受惩罚制约,对他人合作的信任就会被破坏,因此需要适当保持惩罚制度以确保合作。然而, Macy (1993)的情境研究却得出了与本研究相反的结论。他的模拟模型显示,当组织中的合作率较低时,惩罚有助于合作;当组织中的高合作率已经趋于稳定时,惩罚的积极作用消失。此模型预示了组织中已经产生了足够的合作,取消惩罚不会引起合作率的下降。有两点理由可以对此结论进行质疑。首先, Macy 的模型并没有考虑合作的归因和惩罚对人际信任可能的影响。如果考虑到这些方面,他的模型可能会预测到惩罚的缺失导致博弈者人际信任水平的下降。因此,惩罚的贡献可能会持续。其次,在 Macy 的模型中惩罚制度从来没有真正地被取消过。尽管当合作保持在一个较高水平时,惩罚的贡献下降。但惩罚仍然有机会发挥积极作用。所以,这个模型描述了这样一种情形:在合作达到高水平之后,惩罚被运用的机会少了,并非惩罚产生作用的机会完全消除。近来的研究已经表明,一旦惩罚制度制定,人们就不想放弃它(Mulder, Van Dijk, De Cremer, & Wilke, 2005)。

4.3 惩罚、人际信任和合作

在实验第二阶段,经历了惩罚的亲社会价值取向的被试的人际信任水平出现了显著下降,而经历了同样惩罚的个体价值取向的被试的人际信任水平无显著变化,这一结果证实了实验二的预期。

就惩罚对人际信任的削弱效应是否也影响了博弈者的合作行为这一问题,本研究也可以进行解释说明。在实验一中,研究并未将被试最初的人际信任倾向考虑在内,实验一的结果表明惩罚并没有破坏博弈者的合作程度。对阶段2被试的人际信任水平和合作程度进行相关分析,尽管结果表明相关,但这种相关程度似乎并不足以涉及惩罚对合作的

影响。实验二的结果对这一问题进行了有效阐释: 实验二通过引入社会价值取向这一变量区分了被试初始的人际信任水平。结果显示, 惩罚的确破坏了博弈者的合作行为, 但仅限于亲社会价值取向的被试。研究所做的中介效应分析支持了研究预期: 惩罚对亲社会型被试的合作程度有消极影响, 并且这种消极影响通过被削弱的人际信任水平这一中介变量发挥作用。因此, 惩罚似乎已经引起了被试人际信任水平的下降, 这一点在惩罚取消后显现出来。当被试最初的信任水平较高时, 这种被削弱的人际信任似乎已经引起合作程度的下降。在实验一中, 研究并没有测量或是操作初始的人际信任水平, 可能仅仅是初始信任水平较高的被试受信任水平被削弱的影响进行了行为决策, 惩罚对合作的影响也仅仅发生在初始信任水平较高的被试身上, 而人际信任倾向较低的被试则没有受到影响。

惩罚对不同社会价值取向的被试产生不同效应, 一种可能的解释是对于亲社会型被试来说, 惩罚提供的信息与他们的预期不一致, 即惩罚会成为他们降低甚至是不信任同伴的认知线索, 而他们初始的人际信任程度却很高, 所以两者不一致。这种解释受启发于 Pillutla 和 Chen (1999) 的研究, 研究显示如果在社会两难情境中的被试了解到他人的行为与自己的预期不一致, 这将会影响博弈者的博弈行为。然而, 如果他人的行为与自己的预期一致, 则不会影响博弈者的博弈行为。在本研究中, 惩罚尤其使得亲社会型被试对同伴行为的预期和初始的预期不一致。Olson 和 Janes (2002) 的“同伴是不值得信任的”这一认知观念可能会引起被试的惊奇、失望、愤怒和警惕。这种对被试认知的消极唤醒可能驱使人们在进行行为选择时的顾虑, 从而导致被试较低的合作水平。然而, 在本研究中, 惩罚对个体价值取向被试的人际信任水平无显著影响。

Tenbrunsel 和 Messick (1999) 的研究表明, 制裁系统带来知觉决策框架(perceived decisional frame) 的负面效应。两难决策常常是与伦理道德息息相关的, 假如决策时已经知道存在某种制裁系统, 人们就会认为决策是与交易联系在一起的, 是一种市场交换, 更多的精于计算而忽略了决策的道德层面, 即制裁系统改变了人们的合作动机, 从道德的、伦理的动机变成计算的、交易的动机。由于制裁系统是强制性的, 强调人们必须合作, 而不是自愿选择

合作, 它剥夺了人们自由做出选择的权利, 破坏了合作愿望, 容易导致心理阻抗, 从而拒绝合作。Fehr (2002) 的研究提出, 主流的观点认为人是自私自利的, 忽略了建立在公平基础上的利他主义也是合作的重要原因之一, 存在的制裁系统对人们的利他行为产生有害的影响。Mulder 等人(2005, 2006) 的研究把人们的合作动机分为两种: 内在的合作动机和外在的合作动机, 使用“取消制裁”的实验范式验证了制裁对内在的合作动机和信任有破坏的效果, 认为把合作和信任完全建立在制裁基础上有失偏颇, 制裁虽然增加了外在的合作动机, 但减少了人际信任, 一旦取消制裁, 它对合作的破坏就表现出来。惩罚对博弈者人际信任水平的削弱作用之所以在取消惩罚后表现出来, 依据 Mulder 等人的研究结论, 惩罚的存在使得博弈者基于惩罚产生条件信任, 所以当惩罚存在时, 博弈者对他人合作的信任程度的变化无法将条件信任的增加与人际信任的减少进行有效剥离。此外, 对于高人际信任倾向的人来说, 惩罚可能是组织成员不信任他人的认知线索, 他们可能认为组织内部需要惩罚来强制合作。因而, 随着惩罚的取消, 这种建立在惩罚基础之上的制度信任会随着惩罚的取消而消失, 而对于高人际信任倾向的博弈者而言, 会认为没有惩罚制度的约束, 其它团队成员不会自觉合作, 为了确保自己的利益不受损失便不会合作。

5 结论

(1) 在公共物品两难中, 惩罚削弱博弈者的人际信任水平, 惩罚取消后, 经历过惩罚的被试的人际信任水平显著低于没有经历过惩罚的被试。

(2) 惩罚对博弈者合作行为的消极影响在社会价值取向不同的博弈者之间有差异。惩罚取消后, 经历过惩罚的亲社会价值取向被试的合作程度显著下降, 而个体取向被试的合作程度则无显著变化。

(3) 经中介效应检验发现, 人际信任在惩罚影响亲社会价值取向被试的合作程度中起中介作用。惩罚通过亲社会型博弈者的人际信任水平对合作程度有间接负效应, 即惩罚程度越强, 在实验阶段 2, 亲社会型博弈者的人际信任水平越低, 而此时亲社会型博弈者的人际信任水平会影响博弈者的合作程度, 使其合作程度也下降。

参 考 文 献

- Blomqvist, K. (1997). The many faces of trust. *Scandinavian Journal of Management*, 13, 271–286.
- Caldwell, M. D. (1976). Communication and sex effects in a five-person prisoner's dilemma game. *Journal of Personality and Social Psychology*, 33, 273–280.
- Dawes, K. M. (1980). Social Dilemmas. *Annual Review of Psychology*, 31, 169–193.
- Chen, X., & Ye, H. S. (2007). Review of the research in cooperation in social dilemmas. *Advances in Psychological Science*, 15(5), 743–748.
- [陈欣, 叶浩生. (2007). 两难中合作行为研究的回顾和展望. *心理科学进展*, 15(5), 743–748.]
- De Cremer, D., Dewitte S. S., & Snyder, M. (2001). 'The less I trust, the less I contribute? 'The effects of trust, accountability and self-monitoring in social dilemmas. *European Journal of Social Psychology*, 31, 953–105.
- De Cremer, D., & Van Dijk, E. (2002). Reactions to group success and failure as a function of identification level: A test of the goal-trans-formation hypothesis in social dilemmas. *Journal of Experimental Social Psychology*, 38, 435–440.
- De Dreu, C. K. W., Giebels, E., & Van de Vliert, E. (1998). Social motives and trust in integrative negotiation: The disruptive effects of punitive capability. *Journal of Applied Psychology*, 83, 408–422.
- De Dreu, C. K. W., McCusker, C. (1997). Gain-loss frames and cooperation in two person social dilemmas : a transformation analysis. *Journal of Personality and Social Psychology*, 72, 1093–1106.
- Eek, D., Loukopoulos, P., Fujii, S., & Garling, T. (2002). Spill-over effects of intermittent costs for detection in social dilemmas. *European Journal of Social Psychology*, 32, 801–813.
- Fehr, E., & Falk, A. (2001). Psychological foundations of incentives. *European Economic Review*, 46, 670–723.
- Gneezy, U., & Rustichini, A. (2000). A fine is a price. *Journal of Legal Studies*, 29, 8–19.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137–140.
- Huang, Z. B., Wan, D. F. (2005). Cost Advantages of Punishments in Long Term Incentive. *Management Sciences in China*, 18(5), 17–20.
- [黄湛冰, 万迪昉. (2005). 惩罚在长期激励中的成本优势分析. *管理科学*, 18(5), 17–20.]
- Kollock, P. (1998). Social Dilemmas: The Anatomy of cooperation. *Annual Review of psychology*, 24, 183–214.
- Kopelman, S., Weber, J. M., & Messick, D. M. (2002). Factors influencing cooperation in commons dilemmas: A review of experimental psychological research. In E. Ostrom, T. Dietz, P. C. Dolsak, P. C. Stern, S. Stonich, & E. U. Weber (Eds.), *The drama of the commons* (pp.109–160). Washington, DC: National Academy Press.
- Li, S. (2005). Choice reversals across certainty, uncertainty and risk: the equate-to-differentiate interpretation. *Acta Psychologica Sinica*, 37, 427–433.
- [李纾. (2005). 确定、不确定及风险状态下选择反转:“齐当别”选择方式的解释. *心理学报*, 37, 427–433.]
- Li, S. (2006). One Year of Progress in Research on Behavioral Decision-Making. *Advance in Psychological Science*, 14(4), 490–496.
- [李纾. (2006). 发展中的行为决策研究. *心理科学进展*, 14(4), 490–496.]
- Liebrand, W. B. G., Jansen, R. W. T. L., Rijken, V. M. (1986). Might over morality: Social values and the perception of other players in experimental games. *Journal of Experimental Social Psychology*, 22, 203–205
- Liebrand, W. B. G., Messick, D. M., & Wilke, H. A. M. (Eds.). (1992). *Social dilemmas: Theoretical issues and research findings*. Oxford: Pergamon Press.
- Liu, C. J., Li, Y. M., & Li, S. (2007). Advance in research on social dilemmas: Evidence from experimental social psychology. *Advance in Psychological Science*, 15, 379–384.
- [刘长江, 李岩梅, 李纾. (2007). 实验社会心理学中的社会困境. *心理科学进展*, 15, 379–384.]
- Lubmann, N. (1988). Familiarity, confidence, trust: Problems and alternatives. In D. Gambetta (Ed.), *Trust making and breaking Relationships* (pp. 1–18). Oxford: Basil Blackwell.
- Macy, M. W. (1993). Backward-looking social control. *American Sociological Review*, 58, 819–836.
- Mayer, R. C, Davis, J. H., Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20, 709–734.
- McCusker, C., & Carnevale, P. J. (1995). Framing in resource dilemmas: Loss aversion and the moderating effects of sanctions. *Organizational Behavior and Human Decision Processes*, 61, 195–200.
- Messick, D. M., & Brewer, M. B. (1983). Solving social dilemmas: A review. In L. Wheeler & P. Shaver (Eds.), *Review of Personality and Social Psychology* (pp. 35–42). Beverly Hills, CA: Sage.
- Mulder, L. B., Van Dijk, E., De Cremer, D., & Wilke, H. A. M. (2005). When sanctions fail to increase cooperation in social dilemmas: Considering the presence of an alternative defection option. *Personality and Social Psychology Bulletin*, 32, 1312–1324.
- Mulder, L. B., Van Dijk, E., & Wilke, H. A. M. (2006). Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas. *Journal of Experimental Social Psychology*, 42, 147–162.
- Olson, J. M., & Janes, L. M. (2002). Asymmetrical impact: Vigilance for differences and self-relevant stimuli. *European Journal of Social Psychology*, 32, 383–393
- Ostrom, E., Walker J., & Gardner R. (1992). Covenants with and without a sword: self-governance is possible. *American Political Science Review*, 86, 404–417.
- Pillutla, M. M., & Chen, X. P. (1999). Social norms and cooperation in social dilemmas: The effect of context and feedback. *Organizational Behavior and Human Decision Processes*, 78, 81–103.
- Robbins, T. L. (1995). Social loafing on cognitive tasks: An examination of the “sucker effect”. *Journal of Business and psychology*, 9, 340–342.
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. C. (1998). Not so different after all: a cross-discipline view of trust. *Academy of Management Review*, 23, 393–404.
- Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic “remedies” for trust/distrust. *Organization Science*, 4, 367–392.

- Tenbrunsel, A. E. (1999). Trust as an obstacle in environmental-economic disputes. *American Behavioral Scientist*, 42, 1350–1367.
- Tenbrunsel, A. E., & Messick, D. M. (1999). Sanctioning systems, decision frames, and cooperation. *Administrative Science Quarterly*, 44, 684–707.
- Van Vugt, M., & De Cremer, D. (1999). Leadership in social dilemmas: The effects of group identification on collective actions to provide public goods. *Journal of Personality and Social Psychology*, 76, 590–598.
- Wen, Z. L., Zhang, L., Hou, J. T., & Liu, H. Y. (2004). Testing and application of mediating effects. *Acta Psychologica Sinica*, 36, 614–620.
- [温忠麟, 张雷, 侯杰泰, 刘红云. (2004). 中介效应检验程序及应用. *心理学报*, 36, 614–620.]
- Wit, A., & Wilke, H. A. (1990). The presentation of rewards and punishments in a simulated social dilemma. *Social Behavior*, 5, 235–242.
- Yamagishi, T. (1986a). The structure/goal expectation theory of cooperation in social dilemmas. *Advance in Group Process*, 3, 51–87.
- Yamagishi, T. (1986b). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51, 110–116.
- Yamagishi, T. (1988a). The provision of a sanction system in the United States and Japan. *Journal of Experimental Social Psychology Quarterly*, 51, 265–271.
- Yamagishi, T. (1988b). Seriousness of social dilemmas and provision of a sanction system. *Social Psychology Quarterly*, 51, 32–42.
- Yamagishi, T. (1992). Group size and the provision of a sanctioning system in a social dilemma. *Journal of Personality and Social Psychology*, 77, 337–349.
- Yan, J., & Wang, C. M. (2000). The effects of values orientation on group cooperation in social dilemmas. *Acta Psychologica Sinica*, 32, 332–336.
- [严进, 王重鸣. (2000). 两难对策中价值取向对群体合作行为的影响. *心理学报*, 32, 332–336.]
- Yan, J., & Wang, C. M. (2002). The utility transformation of task structure and values in social dilemmas. *Acta Psychologica Sinica*, 34, 529–533.
- [严进, 王重鸣. (2002). 两难情景下任务结构与价值取向的效用特征转换. *心理学报*, 34, 529–533.]
- Yan, J., & Wang, C. M. (2003). The development of cooperative behaviors in resource dilemmas. *Acta Psychologica Sinica*, 35, 499–503.
- [严进, 王重鸣. (2003). 群体任务中合作行为的跨阶段演变. *心理学报*, 35, 499–503.]
- Zhang, C. X. (2006). Research on the relationship of credit loss and punishment. *Market Modernization*, 470, 311–312.
- [张朝孝. (2006). 非重复博弈条件下信誉缺失与惩罚的关系及博弈模型研究. *商场现代化*, 470, 311–312.]

The Effects of Sanction and Social Value Orientation on Trust and Cooperation in Public Goods Dilemmas

WANG Pei; CHEN Li

(School of Education, Shanghai Normal University, Shanghai 200234, China)

Abstract

Sanctioning systems in social dilemmas are often discussed to increase trust in others and to increase cooperation. Yamagishi (1986a, 1986b, 1992) showed that people tended to favor the introduction of a sanctioning system when there was little trust that other people would cooperate. The trust positively relates to cooperation (De Cremer et al., 2001). This suggests a sanctioning system may increase cooperation by increasing trust in others. However, findings outside social dilemma research argued that a sanctioning system can decrease trust that others are internally motivated to cooperate (De Dreu et al, 1998). If a sanctioning system decreased trust in others, would the sanctioning negatively affect cooperation with others? Given this, this paper is to explore the relationship of sanction, trust and cooperation by “Removing the Sanction” paradigm and a measurement of social value orientation.

This paper examined the effects of a sanction on the trust and cooperation in public goods dilemma, including two experiments. In the first experiment, 76 participants were randomly assigned to the two conditions. The level of the trust and cooperation for the participants in the two conditions (sanction versus no-sanction)

were compared after removing the sanction. 126 participants were measured by social value orientation, and 96 participants were chosen including 48 pro-socials and 48 individualists who were randomly assigned to three conditions including no sanction, mild sanction and severe sanction in the second experiment. After removing the mild sanction and severe sanction respectively, the paper compared the trust and cooperation of pro-socials with participants who had not gone through sanction. The same procedure had been done with individualists.

The main conclusions are as follows: participants who have experienced the presence of the sanction trust that another group member would cooperate less than participants who have not. The former cooperates less than the latter. The negative effect of the former presence of the sanction on cooperation is mediated by the trust in phase 2.

This study has important implications in the practice of organization management. In some organizations, when the trust in others is of initially high degree, a sanction installing can undermine this trust and consequently undermine cooperation after removing the sanction. When the degree of trust in others is initially low, it's essential to install a sanction in order to increase cooperation with each other.

Key words public goods dilemmas; sanction; trust; social value orientation; cooperation