

奖惩机制对两难困境中合作行为的影响模式及其差异

周靖昆 崔丽莹*

(上海师范大学心理学院, 上海 200234)

摘要:作为一种重要的激励机制,奖惩机制是通过奖励合作者和惩罚搭便车者促进社会困境中的合作,但二者对合作的影响过程或模式存在差异。一方面,奖惩机制对合作的影响存在效果差异,奖惩机制如何发挥出最优效果会受到不同奖惩环境的影响;另一方面,奖惩机制对合作的影响存在过程差异,奖惩机制会通过诱发互惠机制、情绪机制、行为机制、生理机制来促进合作;同时奖惩机制也会通过诱发资源损耗、认知损耗、行为损耗来破坏合作。未来研究可以扩展考察奖惩对更多类型或情境下合作行为的影响,增加研究结论的生态效度,进一步厘清奖惩对合作影响的最优力度与团体规模,有效识别和确定奖惩机制在促进与破坏合作之间的边界。

关键词:奖励;惩罚;合作;社会困境;边界

中图分类号:B848

文献标志码:A

文章编号:1003–5184(2025)02–0143–10

1 引言

合作是人类赖以生存的重要手段,是社会组织、动物群体、以及物种间关系完整性的重要组成部分(Flores et al., 2021)。作为亲社会行为的重要表现形式,合作是个体或群体之间为了共同的目标而付出自己的努力,既有利于自己,又有利于他人的意向或行为。在利益冲突情境中,合作也是成本和收益权衡的结果(李晶,朱莉琪,2014;Henrich & Henrich, 2006)。目前对合作行为的实验研究主要采用两难博弈困境范式,包含两人困境(如囚徒困境)和多人困境(如公共物品困境和资源困境),考察人们在个人利益与集体利益冲突时的选择。虽然合作强调互惠和共赢,但并非每个人都会选择合作。一方面,合作代价可能很高;另一方面,短期内搭便车者比贡献者获益更多。在这种情况下,合作就会消失,出现公地悲剧(Rankin, Bargum, & Kokko, 2007)。此时奖惩机制作为一种重要的激励机制,可以有效解决搭便车问题,促进合作发展(Balliet, Mulder, & Van Lange, 2011; Wu, Luan, & Raihani, 2022)。

研究发现,无论是哪种激励,只要激励金额大于合作者与叛逃者之间的收益差,就会发生合作行为(Wang, Chen, & Szolnoki, 2019)。奖励是一种正激励,通过对合作者施加社会奖励来增加合作者的收入,从而促进合作意向;惩罚是一种负激励,通过制裁那些违背社会准则的个体,加强社会规范行为。由于采取的激励内容与方式不同,奖励与惩罚对合作的影响效果、过程和模式都存在着差异。那么,两

者之间的具体差异到底体现在哪些方面?这些差异的产生或形成机制是什么?针对目前在奖惩功能方面存在的一些分歧和争议,本文将从两难困境中奖惩对合作的影响着手,着重分析和阐释二者对合作影响效果差异以及影响过程的差异。

2 奖惩机制在合作促进中的影响效能争议

奖励与惩罚是促进合作的重要机制,但两者各有优劣,难于抉择。“大棒”可以帮助人们远离自私行为,“胡萝卜”则可以鼓励人们用合作来取代背叛。但目前学者对于“胡萝卜”与“大棒”谁的效能更佳,仍存在争议(Balliet, Mulder, & Van Lange, 2011; Buskens, Corten, & Snijders, 2020)。

2.1 “胡萝卜”与“大棒”谁更胜一筹?

2.1.1 “胡萝卜”优于“大棒”

一些学者认为,奖励的效果要好于惩罚,能促使团队成员贡献更多的公共物品(Heine & Strobel, 2020)。其优势主要体现在三个方面:其一,奖励可以带来积极的情绪体验和稳定的助推作用,有助于促进行为的保持。惩罚可能会对个体产生消极的情绪反应,如害怕、恐惧等,还可能造成总体资源的亏损,不仅对贡献影响较小,而且会导致回报降低(Rand et al., 2009);惩罚不仅无法维持较高合作水平,而且会降低群体的适应性,而奖励在人类合作中具有重要的助推作用,尤其在一个不确定的环境中,奖励制度不仅是有效而且还是高效的(Dong, Sasaki, & Zhang, 2019),可以避免惩罚带来的问题并强化社会结构(Yang et al., 2018)。

* 通信作者:崔丽莹, E-mail: cui720926@163.com。

其二,奖励的优势不仅存在于一阶激励中,即通过贡献和回报的增加来唤起更多的合作,还存在于二阶激励中。研究发现,免费搭乘行为更容易被二阶奖励(奖励那些主动遵循奖惩规范的人)而不是二阶惩罚(惩罚那些不主动实施奖惩规范的人)所遏制(Kiyonari & Barclay, 2008)。

其三,奖励更符合人们的激励偏好和预期。受互惠原则影响(Isoni & Sugden, 2019),激励者可能会更倾向选择奖励而非惩罚,尤其在初始阶段(Sefton, Shupp, & Walker, 2007)。当可以选择时,人们更愿意将资源用来补偿或奖励合作者(Heffner & FeldmanHall, 2019; Molenmaker, de Kwaadsteniet, & van Dijk, 2016),或奖励那些奖励他人的人,实施二阶奖励而不是二阶惩罚(Kiyonari & Barclay, 2008)。同时人们也更愿意被他人奖励,当个体被奖励时会产生一种积极情绪,从而愿意为群体作出更多的贡献(Rai et al., 2018)。

2.1.2 “大棒”优于“胡萝卜”

一些学者认为惩罚效果优于奖励(Buskens, Corten, & Snijders, 2020; Mulder, 2008; Zhang, An, & Dong, 2021),在人类社会中,人们可能更喜欢使用惩罚(Wang, Chen, & Szolnoki, 2019)。惩罚的优势主要体现在以下三个方面:其一,惩罚带有强制性和威胁性。在损失厌恶效应作用下,人们对资源亏损更加敏感,因此惩罚的威胁可能对个体行为产生有效约束。Mulder(2008)认为惩罚传达了一种义务规则(具有强制性),而奖励传达的是一种自愿规则,因此相比于奖励而言惩罚对个体的影响更大(Mulder, 2008)。近期有学者发现,与最优奖励机制比,最优惩罚机制是提升合作水平的更便捷和便宜的一种方式(Wang, Chen, & Szolnoki, 2019);在促进群体间知识共享方面,惩罚效果优于奖励(Zhang, Song, & Song, 2020)。奖励可能导致合作者和叛逃者的稳定共存,而惩罚可有效遏制合作群体中的自私行为,且中等程度的惩罚要比中等程度奖励更有利于促进合作(Zhang, An, & Dong, 2021)。

其二,惩罚的维持成本低,且具有“增值效应”(multiplication effect),即“大棒”具有“增值性”。奖励机制通常在奖励者和受助者之间的资源水平不对等的情况下才有意义(Hilbe & Sigmund, 2010),并且奖励需要在每次完成合作任务后都要进行。然而当每个成员都选择合作时,惩罚就不再被需要,但产生的威慑作用仍然存在,所以胡萝卜比大棒在维持合作上显得更昂贵(Dari - Mattiacci & De Geest, 2009)。

其三,“大棒”对自利者更适用。奖励不会轻易改变自利者的选择,然而由于惩罚的代价是昂贵的,所以当自利者被惩罚后收益受损,此时回归合作才是最优选择(Buskens, Corten, & Snijders, 2020)。因此,惩罚可以对合作产生稳定持久的作用,而奖励并未解决搭便车的行为,不能有效稳定合作(Hauert, 2010; Hilbe & Sigmund, 2010)。

2.2 “胡萝卜”与“大棒”的协同性

近年来越来越多的研究者跳出奖励和惩罚非此即彼的抉择,将目光转向奖惩激励的相互作用,关注奖惩的协同效应对合作的影响(Chen et al., 2015; Liu et al., 2019)。作为一种双向激励,奖惩协同不仅正向激励合作者,而且负向激励叛逃者,比单独奖惩的效果更好,协同出现越频繁,对合作的促进越好(Fang et al., 2019)。

然而,奖惩协同也会增加成本,激励者既要承担奖励成本也要承担惩罚成本,最终可能影响整体收益。那么,应该如何更好发挥奖惩机制的协同作用呢?首先,通过调整奖惩的顺序来促进合作,即先用“胡萝卜”再用“大棒”,先用奖励来稳定合作者,当合作者的合作频率超过基线时,再将激励从奖励转变为惩罚。与单独奖惩机制相比,该方式所产生的成本低且可以广泛的促进合作,最大限度地减少叛逃者的优势(Chen et al., 2015; Hilbe & Sigmund, 2010)。其次,需要考虑所处的博弈环境,在单次或博弈初期时惩罚可能更有效,而在后期的反复博弈中奖励可能更有效。在单次或初期博弈中,奖励虽然激励了合作者,但无法维持合作,更适合采用惩罚(Rand et al., 2009; Sefton, Shupp, & Walker, 2007)。因为选择背叛获得的收益要比选择合作带来的更大,尤其在在博弈轮数较少时,对参与者来说重要的是实现收益最大化,而是否获得奖励并不重要,即不投资可能是主导策略(van den Berg et al., 2020)。但在多轮重复博弈时,奖励则优于惩罚,因为从整体来看此时奖励可使群体最终的获益更高(Rand et al., 2009)。

3 奖惩机制在合作促进中的影响过程差异

虽然奖惩都可以促进合作,但奖励带来“获得”,而惩罚带来“损失”,两者对合作的促进过程或模式必然存在不同。文献分析发现,奖惩机制的差异具体表现在互惠、情绪、行为和生理等方面。

3.1 互惠机制:积极互惠 vs 消极互惠

互惠是维持合作的重要因素,是个体将自己的行为与伙伴的行为调整一致的过程。互惠可以通过奖励合作者和惩罚叛逃者来促进合作(Salazar et

al., 2022)。互惠包括积极互惠和消极互惠,前者是以善报善,后者是以恶报恶。在互惠机制中,奖励往往通过积极互惠来促进合作,而惩罚往往通过消极互惠来促进合作。

积极互惠是对积极行为的回报,可以促进合作的发展 (Pei, Yan, & Wang, 2021; Wu, Zhang, & Chang, 2019), 其原因是积极互惠强调对积极行为施加正面奖励,促使成员之间形成彼此互帮互助,相互激励的氛围,不仅增加了群体收益,还使成员之间形成了良好的合作互惠关系 (Wu, Zhang, & Chang, 2019)。在积极互惠条件下,彼此之间的互动奖励可以让合作群体更加繁荣 (Rand et al., 2009)。

消极互惠是对消极行为的回报,可以改变被惩罚的负向行为,推动合作的发展 (Fehr & Gächter, 2000; Friedman & Singh, 2004)。当有叛逃者的出现时,会引发消极互惠者的愤怒情绪,此时会通过惩罚叛逃者的方式来消解 (Harth & Regner, 2017; Wu, Luan, & Raihani, 2022)。消极互惠的出现意味着被惩罚者的行为不符合社会规范的要求,此时可以促使被惩罚者通过改变自身行为来符合社会规范。消极互惠对合作的促进作用可以用惩罚的社会规范激活作用来解释。惩罚通过引发的描述性规范 (injunctive norms)——“大多数人怎么做”和命令性规范 (descriptive norms)——“应该怎么做”来矫正被惩罚者的不合规行为 (陈思静, 何铨, 马剑虹, 2015)。在两难博弈中,成员会通过利他惩罚来回报复逃者的负向行为 (Greenwood, Abbass, & Petraki, 2018), 并且还可以通过溢出效应来影响叛逃者后期在其他博弈环境中的行为 (陈思静, 何铨, 马剑虹, 2015; 陈思静等, 2021)。

3.2 情绪机制:感恩 vs 内疚、羞耻感

奖惩是影响情绪的重要因素,当个体获得奖励时会唤起正向情绪,受到惩罚时会唤起负向情绪。其中情绪通过诱发异质性投资 (面对合作邻居或叛逃者时情绪变化所引发的投资行为) 可以显著预测合作水平 (Long, Gong, & Yao, 2022)。奖励可能诱发个体的感恩来促进合作,而惩罚可能诱发个体的内疚或羞耻感来促进合作。

感恩是一种积极的情感和特质 (Karns, Moore, & Mayr, 2017)。当个体受到他人慷慨利他行为时会产生感激之类的积极情绪 (McCullough et al., 2001), 这种情绪又会激活个体内在的亲社会动机 (Balconi & Fronda, 2021), 促使个体产生亲社会行为,维持施惠者与受惠者之间持续的互惠关系 (Shiraki & Igarashi, 2018)。研究发现感恩的人会对

他们的恩人做出更多的社会包容行为,即使这些行为可能会让自己利益受损 (Bartlett et al., 2012)。其中高感恩水平的个体更容易在社会困境中表现出合作行为 (任海霞等, 2021; DeSteno et al., 2010)。这意味着,在两难博弈中参与者获得奖励时会产生感激之情,为了报恩从而引发出更多的合作行为。

内疚和羞耻是通过自我评价而产生的一种有关自我意识的“道德”情绪,往往由行为的消极结果导致 (Tangney, Stuewig, & Mashek, 2007)。其中,内疚是指当个体做错事之后所产生的悔恨、自责、痛苦的情绪体验;羞耻是指个体消极行为后,意识到不受欢迎或不被认可时所产生的无价值、无能为力的感觉 (Stuewig et al., 2015)。在社会困境中,受到惩罚的叛逃者察觉到他人对自己行为的不满,可能进行消极的内部归因,诱发内疚或羞耻感 (郝娜, 崔丽莹, 2022), 为了补偿他人或保护自尊,这两种情绪都可能增加个体的合作行为 (Nunney, Schalk, & Manstead, 2022)。

3.3 行为机制:增强效应 VS 抑制效应

操作性条件反射揭示了奖励与惩罚对行为影响的内在心理机制 (张大俭, 2016)。奖励可以激发个体产生更多目标行为 (Wu et al., 2017)。小白鼠通过自主按压杠杆后获得食物 (强化), 并在强化刺激的影响下,不断重复这种操作行为。在两难博弈合作中,奖励是一种正向的强化刺激,当合作者得到奖励时会产生更多合作行为,以便可以继续获得奖励。

惩罚可以通过抑制个人负向行为来促使正向行为 (Li, Zhu, & Chen, 2018; Raihani & Bshary, 2019)。当小白鼠按压杠杆时给予电击 (抑制), 它就会减少按压频率,直到按压行为消失。在两难博弈合作中,惩罚作为一种抑制个体叛逃行为的机制,可以有效抑制自利行为。惩罚的抑制性一方面源于惩罚可能带来高昂的资源亏损,个体为减少亏损最终选择回归合作;另一方面在损失规避的影响下,与获得相比,人们对损失更加敏感 (Balcombe et al., 2019; Yu et al., 2021)。

3.4 生理机制:催产素、血清素 vs 海马体、杏仁核

奖惩通过激活个体的生理机制来促进合作的产生。其中神经递质 (催产素、血清素) 与边缘系统 (海马体、杏仁核) 都是负责调节的社会行为的重要生理机制 (Attaran et al., 2019; Rolls, 2019)。奖励对合作的影响主要与催产素、血清素有关,而惩罚对合作的影响主要与海马体、杏仁核有关。

催产素对奖赏具有敏感性 (Bradley et al., 2020; Lee et al., 2014)。主要通过增强壳核、岛叶和

前扣带皮层中奖励和损失预期的神经反应增加了个体对金钱刺激反应。其中壳核反应增强可能会增加个体对奖励的预期并促进个体产生目标行为,前扣带皮层反应增强可能会促使个体追求奖励最大化(Nawijn et al., 2016)。除催产素之外,血清素(5-HT)在促进合作方面也发挥着重要作用(Paula et al., 2015; Pimentel et al., 2019)。其中奖励机制对5-HT起了重要的激活作用,不论是预期还是意外奖励,都会迅速激活5-HT神经元(Li et al., 2016; Nagai et al., 2020)。5-HT通过与中脑边缘多巴胺系统的之间相互作用来调节反应行为(Daw, Kakade, & Dayan, 2002),其中中脑边缘多巴胺活动增强会加强奖励行为出现(Baik, 2020)。这意味着,奖励会激活5-HT神经元,而5-HT会增加会诱发个体的合作行为。

“一朝被蛇咬,十年怕井绳”。普遍恐惧的心理与杏仁核、海马体活动相关,特别是与基底外侧杏仁核(Jean-Richard-Dit-Bressel & McNally, 2015)。白鼠实验发现基底外侧杏仁核的损伤会导致恐惧表达出现障碍,从而导致冒险行为的增加(Orsini et al., 2015)。个体对负性事件的抑制主要是依靠杏仁核和海马体的作用(Hahn et al., 2010),海马体和杏仁核灰质体积的增加与对厌恶事件的敏感性有关,在处理厌恶刺激时表现较为活跃(Barrós-Loscertales et al., 2006),其中对损失的预期还会激活海马体和杏仁核(Hahn et al., 2010)。当搭便车者被惩罚后,在海马体和杏仁核的作用下,搭便车者可能会通过选择合作的方式来规避惩罚。

4 奖惩机制对合作抑制过程的影响差异

由于奖惩机制对合作的影响效果会受到激励方式和博弈环境等因素的制约,因此若应用不当可能会带来负向结果。分析发现奖惩机制可能会通过诱发资源损耗、认知损耗、行为损耗来破坏合作。

4.1 资源损耗:成本损耗 VS 收益损耗

从个体角度来看,奖励是一种“获得”方式,而惩罚是一种“破坏”方式。在合作博弈过程中,惩罚可能导致人们的最终收益下降,而奖励可能使人们的最终收益提高。从激励对象角度来看,奖励带来的资源风险更小。但是从整个群体的角度来看,奖惩在资源损耗上的影响又会随着时间变化而变化。

首先,从前期的成本损耗来看,由于奖励可能会产生过多的成本,因此在长期合作中可能很难维持下去,从而不利于合作的进行。然而基于惩罚的“增值效应”和“溢出效应”,即惩罚会抑制叛逃者在新博弈情境下的自私行为(陈思静等, 2021),随着

惩罚轮数的增加,惩罚的成本反而在后期可能逐渐减少(Gächter, Renner, & Sefton, 2008)。因此,当资源有限时,惩罚可能是节省成本的不错选择;当资源相对充足时,奖励可能是一种提高群体收益的不错选择。同时,奖惩成本的损耗与奖惩对象的数量有关。在多人博弈中,如果合作者占多数,那么奖励自然是不划算的;相反,如果博弈中叛逃者占多数时,那么惩罚就变得不划算(Hilbe & Sigmund, 2010; Hauert, 2010)。

其次,从后期的收益损耗来看,持续互动后惩罚的损耗可能更高,收益可能更少,从而可能会降低成员合作的意愿。同时不同的激励类型也会影响最终收益。从惩罚来看,间接激励的收益效果较好;而从奖励来看,直接激励的收益效果较好。研究发现,一方面,直接惩罚对个体来说是昂贵的,尤其代价高昂的直接惩罚更不利于群体间的反复互动(Ule et al., 2009);另一方面,直接惩罚更容易受到以牙还牙的“回报”,导致合作资源的损耗,而使用间接惩罚策略的风险较小(Balafoutas, Nikiforakis, & Rockenbach, 2014; Molho et al., 2020)。与惩罚不同,直接互惠奖励可以更好促进合作发展(Güth et al., 2000; Pei, Yan, & Wang, 2021; Wu, Zhang, & Chang, 2019)。一方面直接互惠奖励中个体更容易信任他人,且容易加入到这种互惠“契约”中,而间接互惠奖励缺乏这种“契约”感;另一方面直接互惠通过社会比较,更可能促进个体的贡献率(Güth et al., 2000)。

4.2 认知损耗:转移注意 VS 破坏信任

许多研究发现,奖惩机制在一定程度上会破坏个体内在亲社会动机,并出现对“内在动机”的挤出现象(Chao, 2017; Gneezy, Meier, & Rey-Biel, 2011),这一过程可能是通过转移个体的注意和破坏成员之间的信任而实现的。

奖励机制的出现可能会使得被激励者注意力发生变化,使他们关注点更倾向于奖励本身而非亲社会行为,导致以亲社会为目标转变为以关注成本效益为目标,最终降低了利他行为(Bénabou & Tirole, 2006)。例如,在慈善捐赠活动中的感恩奖励(例如杯子)可能会导致事与愿违的结果—降低捐款率(Chao, 2017; Newman & Jeremy Shen, 2012)。这意味着在注意机制驱动的影响下,参与者可能会衡量投入与收益之间的比值,从而只贡献适量的资源,以防止过多的贡献导致亏损。

与奖励不同,惩罚可能更容易破坏个体对他人的信任(Wang & Murnighan, 2017)。虽然有利于维

护公平正义,但惩罚也会营造一种残酷的氛围,一方面让惩罚者树立一种严苛形象,另一方面也减少成员之间的情感联结。在群体中惩罚者往往是不受欢迎并且还可能不被信任,尤其惩罚行为涉及到自身利益时(Kiyonari & Barclay, 2008; Strimling & Eriksson, 2014; Wang & Murnighan, 2017)。当个体对他人的信任水平降低时,通常会减少贡献来维护自己的利益,从而导致合作水平下降(Balliet & Van Lange, 2013a)。低信任水平个体更容易对惩罚做出消极回应,更容易使合作水平降低(Balliet & Van Lange, 2013b)。除此之外,惩罚体系的存在可能会让人们认为他人的合作行为更多是为了自己的利益,从而摧毁合作信念(Mulder et al., 2006)。在惩罚取消之后,个体对成员信任水平会变得更低,从而进一步导致较低的合作贡献率。即使当初信任水平很高个体在经历过制裁体系之后其信任水平和合作率也会下降(王沛,陈莉, 2011; Mulder et al., 2006)。对参与者而言,当制裁体系取消后意味着对叛逃者的制约消失,降低合作贡献率可以防止自己利益受损。

4.3 行为损耗:反社会奖励 vs 反社会惩罚

近年来人们发现,社会困境中的奖惩也存在着行为风险,出现反社会奖励或惩罚。虽然人们总希望通过奖励合作者和惩罚搭便车者(亲社会奖惩)来促进合作,但由于反社会行为在人类社会的存在,合作者也可能受到惩罚,搭便车者受到奖励,尤其在缺乏信息情况下(Bruhin, Janizzi, & Thöni, 2020; Du et al., 2018; dos Santos & Peña, 2017)。当激励体系被用来奖励叛逃者,制裁合作者就是反社会奖惩。一旦反社会激励者的人数增加,搭便车者会形成“叛逃者联盟”,从而破坏了合作(Herrmann, Thöni, & Gächter, 2008; Rand et al., 2010)。反社会激励的出现会破坏奖惩初衷,造成了奖惩资源的浪费,最终使群体的合作水平下降(陈璟等, 2022; dos Santos, 2015)。

反社会奖励的出现更可能源于激励体系本身的优越性所引发的心理博弈。当个体选择搭便车时,获益会高于选择合作的成员,此时对其他叛逃者的奖励可能会通过互惠的方式来回馈自己,从而增加自己的收益,那么对于叛逃者而言收益实现了最大化(Du et al., 2018; dos Santos, 2015);反社会惩罚的出现一方面可能是为了激励合作者增贡献量,另一方面可能属于报复行为(Herrmann, Thöni, & Gächter, 2008),即搭便车者被实施惩罚之后报复惩罚者或其他合作者。在这种恶性循环下,成员之间

互不信任,合作率下降。

5 现存不足与未来展望

大量的实证研究探索了奖惩机制对合作的影响,但在对两者的影响差异及形成机制进行梳理和对比时,发现以往研究存在三个方面的不足。

奖惩机制对合作的影响证据主要来源于囚徒困境和公共物品困境等经济博弈实验,而较少来自于现实生活中团队合作的研究支持,实验情境的单一会降低实验结果的生态效度。与真实的社会合作不同,实验室或者在线的合作主要采用匿名形式,被试之间缺乏言语和肢体的交流,陌生人之间也较少受到声誉、八卦等信息的影响。研究发现,当个体知道自己的信息(例如名字)将会被他人知道时,他们对公共利益的贡献率增加了50%(Jacquet et al., 2012)。因为实名影响下个体会有一种“暴露”在外界的感觉,不合作可能会威胁到个体声誉,引发羞愧感。虽然许多线上实验结果表明奖惩机制可以有效地促进合作,但目前已有研究发现在线下真实互动中的奖惩可能并不能有效的促进合作(Noussair, van Soest, & Stoop, 2015),并且线上的实验结果可能高估了利他惩罚的意愿(Pedersen et al., 2020),真实生活中间接惩罚可能比直接惩罚更加频繁(Balafoutas, Nikiforakis, & Rockenbach, 2014)。同时,在实验情境中,随着互动轮次、任务和对象的变化,奖惩对合作的影响也会发生变化。例如,一次性互动合作中惩罚更有效,同组多轮重复合作中奖励更有效(Rand et al., 2009);惩罚对于不重新分配成员的多轮互动的效果要好于重新分配成员的多轮互动(Balliet, Mulder, & Van Lange, 2011)。那么真实情景中的奖惩是否也会受到互动频率和重复性的影响?未来研究可以考虑加入真实生活中的互动情境,或增加社会困境的类型和轮次,更全面探究奖惩机制的作用。

第二,已有研究表明奖惩力度与团体规模大小会影响奖惩对合作的促进作用(孔程程,王晓明, 2018; Hu et al., 2020),但什么样的“度”对合作影响效果最好,还有待更深入探究。需要厘清的“度”主要包含以下几个:

首先是奖惩的力度。代价高昂的激励虽然可以促进人类之间的合作,但会产生许多弊端。在奖励机制方面,奖励强度越大,金额越高,即使可以充分促进合作,也会导致个人或共同资源的枯竭,相反小金额、强度适中的奖励,可能有利于维持合作与资源的双赢(Hu et al., 2020)。在公共资金总量稳定的情况下,低强度且覆盖范围广的奖励对合作的影响

最大(Li, Chen, & Niu, 2018);在惩罚机制方面,高昂的惩罚降低了继续合作的意愿,当罚金提高时,会导致合作者的整体收益降低,最终导致合作消失(Li et al., 2018; Wu et al., 2009),并且不会产生任何积极的效果,惩罚高昂和没有惩罚的结果可能一样(Helbing et al., 2010)。可见,高昂的奖励和惩罚都是弊大于利,但具体利弊存在差异,未来研究可以进一步比较奖惩适度的最优范围及影响因素。

其次是群体规模的宽度。目前对于团体规模对合作的影响存在争议。一些研究认为小群体规模更有利于合作(Nosenzo, Quercia, & Sefton, 2015; Wheelan, 2009)。因为随着成员增多,个体责任感会降低,容易出现搭便车现象,搭便车现象变得更为隐蔽,此时奖惩机制的作用就减弱(孔程程, 王晓, 2018);一些研究认为大群体规模不一定会导致成员合作下降,尤其当个体收益会随着群体规模增加而增加时(Barcelo & Capraro, 2015; Wu et al., 2020),同时也有研究发现群体规模大小对合作贡献水平并没有显著影响(Zelmer, 2003)。但目前关于奖惩实证研究大多集中在小群体中,因此未来研究可以进一步探索奖惩条件下最优的群体规模。

第三,奖惩机制作为一把双刃剑,既可以促进合作,也会破坏合作(Kitakaji & Ohnuma, 2019; Raihani & Bshary, 2019),那么促进与破坏之间的边界条件是什么?以往研究发现与奖惩的二阶困境和代币化有关。一方面,奖惩机制引发的二阶搭便车和二阶社会困境问题可能会破坏合作。由于奖惩的成本昂贵,可能超过合作的收益,并且激励者可能需要独自承担成本,因此出现了二阶搭便车者,即只贡献合作而不贡献奖惩资源(Fang et al., 2020; Ozono, Kamijo, & Shimizu, 2017)。如果为合作做出贡献但不为激励做出贡献的人收益更高,就会形成二阶社会困境(Ye et al., 2016)。另一方面,奖惩机制的短期和长期影响是否一致往往与奖惩形式有关。以往大部分研究采用金钱激励的方式,即增加合作者或减少叛逃者的代币。研究发现,货币化的奖惩虽然可以增强短期的合作,但是从长期来看可能是有害的。这些研究引入奖惩取消范式,均发现奖惩取消后个体的信任水平受到损害,合作水平下降(王沛, 陈莉, 2011; Mulder et al., 2006; Irwin, Mulder, & Simpson, 2014)。有研究发现,社会批评和社会赞许这种非金钱激励(例如,指责、赞扬)方式也会影响到个体合作行为,甚至其作用要优于金钱激励(陈欣, 赵国祥, 叶浩生, 2014),而且并未发现惩罚取消对合作的破坏(崔丽莹等, 2017)。目前研究更多集中在

奖惩机制对合作的促进作用,对其破坏作用的探究相对较少,未来研究可以尝试探究奖惩机制在合作破坏中的差异,分析谁产生的负面影响更大,以及影响其积极与消极作用转化的边界条件。

参考文献

- 陈璟, 张融, 袁佳琦, 余升翔. (2022). 博弈中的反社会惩罚. *心理科学进展*, 30(2), 436 - 448.
- 陈思静, 何铨, 马剑虹. (2015). 第三方惩罚对合作行为的影响: 基于社会规范激活的解释. *心理学报*, 47(3), 389 - 405.
- 陈思静, 邢懿琳, 翁异静, 黎常. (2021). 第三方惩罚对合作的溢出效应: 基于社会规范的解释. *心理学报*, 53(7), 758 - 772.
- 陈欣, 赵国祥, 叶浩生. (2014). 公共物品困境中惩罚的形式与作用. *心理科学进展*, 22(1), 160 - 170.
- 崔丽莹, 何幸, 罗俊龙, 黄晓娇, 曹玮佳, 陈晓梅. (2017). 道德与关系惩罚对初中生公共物品困境中合作行为的影响. *心理学报*, 49(10), 1322 - 1333.
- 郝娜, 崔丽莹. (2022). 补偿他人还是保护自己? 内疚与羞耻情绪对合作行为的影响差异. *心理科学进展*, 30(7), 1626 - 1636.
- 孔程程, 王晓明. (2018). 惩罚对合作的影响: 形式、机制及边界条件. *心理研究*, 11(2), 166 - 172.
- 李晶, 朱莉琪. (2014). 高功能孤独症儿童的合作行为. *心理学报*, 46(9), 1301 - 1316.
- 任海霞, 黄瑶, 郝娜, 崔丽莹. (2021). 特质感恩与状态感恩对初中生合作水平的影响. *心理科学*, 44(1), 97 - 103.
- 王沛, 陈莉. (2011). 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. *心理学报*, 43(1), 52 - 64.
- 张大俭. (2016). 奖励与惩罚的心理机制及在学校教育中的有效运用. *沧州师范学院学报*, 32(1), 80 - 83.
- Attaran, A., Salahinejad, A., Crane, A. L., Niyogi, S., & Chivers, D. P. (2019). Chronic exposure to dietary selenomethionine dysregulates the genes involved in serotonergic neurotransmission and alters social and antipredator behaviours in zebrafish (*Danio rerio*). *Environmental Pollution*, 246, 837 - 844.
- Baik, J. - H. (2020). Stress and the dopaminergic reward system. *Experimental & Molecular Medicine*, 52(12), 1879 - 1890.
- Balafoutas, L. N., & Rockenbach, B. (2014). Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences*, 111(45), 15924 - 15927.
- Balcombe, K., Bardsley, N., Dadzie, S., & Fraser, I. (2019). Estimating parametric loss aversion with prospect theory: Recognising and dealing with size dependence. *Journal of Economic Behavior & Organization*, 162, 106 - 119.

- Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, 137(4), 594–615.
- Balliet, D., & Van Lange, P. A. M. (2013a). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, 139(5), 1090–1112.
- Balliet, D., & Van Lange, P. A. M. (2013b). Trust, Punishment, and Cooperation Across 18 Societies: A Meta-Analysis. *Perspectives on Psychological Science*, 8(4), 363–379.
- Barrós - Loscertales, A., Meseguer, V., Sanjuán, A., Belloch, V., Parcet, M. A., Torrubia, R., & Ávila, C. (2006). Behavioral Inhibition System activity is associated with increased amygdala and hippocampal gray matter volume: A voxel-based morphometry study. *NeuroImage*, 33(3), 1011–1015.
- Bénabou, R., & Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, 96(5), 1652–1678.
- Barcelo, H., & Capraro, V. (2015). Group size effect on cooperation in one-shot social dilemmas. *Scientific Reports*, 5, 7937.
- Bradley, E. R., Brustkern, J., Coster, L. D., Bos, W., McClure, S. M., Seitz, A., & Woolley, J. D. (2020). Victory is its own reward: Oxytocin increases costly competitive behavior in schizophrenia. *Psychological Medicine*, 50(4), 674–682.
- Bruhin, A., Janizzi, K., & Thöni, C. (2020). Uncovering the heterogeneity behind cross-cultural variation in antisocial punishment. *Journal of Economic Behavior & Organization*, 180, 291–308.
- Buskens, V., Corten, R., & Snijders, C. (Eds.). (2020). *Advances in the Sociology of Trust and Cooperation: Theory, Experiments, and Field Studies*. De Gruyter.
- Bartlett, M. Y., Condon, P., Cruz, J., Baumann, J., & Desteno, D. (2012). Gratitude: Prompting behaviours that build relationships. *Cognition & Emotion*, 26(1), 2–13.
- Balconi, M., & Fronda, G. (2021). Gratitude Affects Inter-Subjective Synchronicity for Cognitive Performance and Automatic Responsiveness. *Frontiers in Psychology*, 12, 574983.
- Chao, M. (2017). Demotivating incentives and motivation crowding out in charitable giving. *Proceedings of the National Academy of Sciences*, 114(28), 7301–7306.
- Cressman, R., Wu, J. -J., Li, C., & Tao, Y. (2013). Game Experiments on Cooperation Through Reward and Punishment. *Biological Theory*, 8(2), 158–166.
- Chen, X., Sasaki, T., Brännström, Å., & Dieckmann, U. (2015). First carrot, then stick: How the adaptive hybridization of incentives promotes cooperation. *Journal of The Royal Society Interface*, 12, 20140935.
- Daw, N. D., Kakade, S., & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks*, 15, 603–616.
- Dong, Y., Sasaki, T., & Zhang, B. (2019). The competitive advantage of institutional reward. *Proceedings of the Royal Society B: Biological Sciences*, 286, 20190001.
- dos Santos, M. (2015). The evolution of anti-social rewarding and its countermeasures in public goods games. *Proceedings of the Royal Society B: Biological Sciences*, 282, 20141994.
- dos Santos, M., & Peña, J. (2017). Antisocial rewarding in structured populations. *Scientific Reports*, 7, 6212.
- Du, C., Jia, D., Jin, L., & Shi, L. (2018). The impact of neutral reward on cooperation in public good game. *The European Physical Journal B*, 91, 234.
- Dari-Mattiacci, G., & Geest, G. D. (2009). Carrots, Sticks, and the Multiplication Effect. *The Journal of Law, Economics, and Organization*, 26(2), 365–384.
- DeSteno, D., Bartlett, M. Y., Baumann, J., Williams, L. A., & Dickens, L. (2010). Gratitude as moral sentiment: Emotion-guided cooperation in economic exchange. *Emotion*, 10(2), 289–293.
- Fang, Y., Benko, T. P., Perc, M., Xu, H., & Tan, Q. (2019). Synergistic third-party rewarding and punishment in the public goods game. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 475, 20190349.
- Fang, Y., Perc, M., & Xu, H. (2020). The Singaporean model in public goods dilemmas with benevolent leaders and bribery. *Journal of Theoretical Biology*, 501, 110345.
- Fehr, E., & Gächter, S. (2000). Fairness and Retaliation: The Economics of Reciprocity. *American Economic Review*, 14(3), 159–181.
- Flores, L. S., Fernandes, H. C. M., Amaral, M. A., & Vainstein, M. H. (2021). Symbiotic behaviour in the public goods game with altruistic punishment. *Journal of Theoretical Biology*, 524, 110737.
- Friedman, D., & Singh, N. (2004). Negative reciprocity: The co-evolution of memes and genes. *Evolution and Human Behavior*, 25(3), 155–173.
- Gächter, S., Renner, E., & Sefton, M. (2008). The Long-Run Benefits of Punishment. *Science*, 322, 1510.
- Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, 25(4), 191–210.
- Greenwood, G. W., Abbass, H. A., & Petraki, E. (2018). When is altruistic punishment useful in social dilemmas? *Biosystems*, 174, 60–62.
- Güth, W., Königstein, M., Marchand, N., & Nehring, K. (2000). *Trust and Reciprocity in the Investment Game with Indirect Reward*. SFB 373 Discussion Paper, No. 2000, 110, Humboldt University of Berlin.
- Hahn, T., Dresler, T., Plichta, M. M., Ehli, A. -C., Ernst, L. H., Markulin, F., Polak, T., Blaimer, M., Deckert, J., Lesch, K. -P., Jakob, P. M., & Fallgatter, A. J. (2010). Functional Amygdala-Hippocampus Connectivity During Antici-

- pation of Aversive Events is Associated with Gray's Trait "Sensitivity to Punishment". *Biological Psychiatry*, 68(5), 459–464.
- Hauert, C. (2010). Replicator dynamics of reward & reputation in public goods games. *Journal of Theoretical Biology*, 267(1), 22–28.
- Heffner, J., & FeldmanHall, O. (2019). Why we don't always punish: Preferences for non-punitive responses to moral violations. *Scientific Reports*, 9, 13219.
- Heine, F., & Strobel, M. (2020). Reward and punishment in a team contest. *PLOS ONE*, 15(9), e0236544.
- Helbing, D., Szolnoki, A., Perc, M., & Szabó, G. (2010). Punish, but not too hard: How costly punishment spreads in the spatial public goods game. *New Journal of Physics*, 12(8), 083005.
- Henrich, J., & Henrich, N. (2006). Culture, evolution and the puzzle of human cooperation. *Cognitive Systems Research*, 7, 220–245.
- Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial Punishment Across Societies. *Science*, 319, 1362–1367.
- Harth, N. S., & Regner, T. (2017). The spiral of distrust: (Non-) cooperation in a repeated trust game is predicted by anger and individual differences in negative reciprocity orientation. *International Journal of Psychology*, 52, 18–25.
- Hilbe, C., & Sigmund, K. (2010). Incentives and opportunism: From the carrot to the stick. *Proceedings of the Royal Society B: Biological Sciences*, 277, 2427–2433.
- Hu, L., He, N., Weng, Q., Chen, X., & Perc, M. (2020). Rewarding endowments lead to a win-win in the evolution of public cooperation and the accumulation of common resources. *Chaos, Solitons & Fractals*, 134, 109694.
- Irwin, K., Mulder, L., & Simpson, B. (2014). The Detrimental Effects of Sanctions on Intragroup Trust: Comparing Punishments and Rewards. *Social Psychology Quarterly*, 77(3), 253–272.
- Isoni, A., & Sugden, R. (2019). Reciprocity and the Paradox of Trust in psychological game theory. *Journal of Economic Behavior & Organization*, 167, 219–227.
- Jean-Richard-Dit-Bressel, P., & McNally, G. P. (2015). The role of the basolateral amygdala in punishment. *Learning & Memory*, 22(2), 128–137.
- Jacquet, J., Hauert, C., Traulsen, A., & Milinski, M. (2012). Could shame and honor save cooperation? *Communicative & Integrative Biology*, 5(2), 209–213.
- Karns, C. M., Moore, W. E., & Mayr, U. (2017). The Cultivation of Pure Altruism via Gratitude: A Functional MRI Study of Change with Gratitude Practice. *Frontiers in Human Neuroscience*, 11, 599.
- Kitakaji, Y., & Ohnuma, S. (2019). The Detrimental Effects of Punishment and Reward on Cooperation in the Industrial Waste Illegal Dumping Game. *Simulation & Gaming*, 50(5), 509–531.
- Kiyonari, T., & Barclay, P. (2008). Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology*, 95(4), 826–842.
- Lee, M. R., Glassman, M., King-Casas, B., Kelly, D. L., Stein, E. A., Schroeder, J., & Salmeron, B. J. (2014). Complexity of oxytocin's effects in a chronic cocaine dependent population. *European Neuropsychopharmacology*, 24(9), 1483–1491.
- Li, J., Zhu, L., & Chen, Z. (2018). The association between punishment and cooperation in children with high-functioning autism. *Journal of Experimental Child Psychology*, 171, 1–13.
- Li, X., Jusup, M., Wang, Z., Li, H., Shi, L., Podobnik, B., Stanley, H. E., Havlin, S., & Boccaletti, S. (2018). Punishment diminishes the benefits of network reciprocity in social dilemma experiments. *Proceedings of the National Academy of Sciences*, 115(1), 30–35.
- Li, Y., Chen, S., & Niu, B. (2018). Reward depending on public funds stimulates cooperation in spatial prisoner's dilemma games. *Chaos, Solitons & Fractals*, 114, 38–45.
- Li, Y., Zhong, W., Wang, D., Feng, Q., Liu, Z., Zhou, J., Jia, C., Hu, F., Zeng, J., Guo, Q., Fu, L., & Luo, M. (2016). Serotonin neurons in the dorsal raphe nucleus encode reward signals. *Nature Communications*, 7, 10503.
- Liu, J., Li, M., Alam, M., Chen, Y., & Wu, T. (2019). A Game Theoretic Reward and Punishment Unwanted Traffic Control Mechanism. *Mobile Networks and Applications*, 24(4), 1279–1294.
- Long, H., Gong, R., & Yao, J. (2022). Heterogeneous investments induced by emotions promote cooperation in public goods games. *Europhysics Letters*, 137(2), 21001.
- McCullough, M. E., Kilpatrick, S. D., Emmons, R. A., & Larson, D. B. (2001). Is gratitude a moral affect? *Psychological Bulletin*, 127(2), 249–266.
- Molenmaker, W. E., de Kwaadsteniet, E. W., & van Dijk, E. (2016). The impact of personal responsibility on the (un)willingness to punish non-cooperation and reward cooperation. *Organizational Behavior and Human Decision Processes*, 134, 1–15.
- Molho, C., Tybur, J. M., Van Lange, P. A. M., & Balliet, D. (2020). Direct and indirect punishment of norm violations in daily life. *Nature Communications*, 11, 3432.
- Mulder, L. B. (2008). The difference between punishments and rewards in fostering moral concerns in social decision making. *Journal of Experimental Social Psychology*, 44(6), 1436–1443.
- Mulder, L. B., van Dijk, E., De Cremer, D., & Wilke, H. A. M.

- (2006). Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas. *Journal of Experimental Social Psychology*, 42(2), 147 – 162.
- Nagai, Y., Takayama, K., Nishitani, N., Andoh, C., Koda, M., Shirakawa, H., Nakagawa, T., Nagayasu, K., Yamanaka, A., & Kaneko, S. (2020). The Role of Dorsal Raphe Serotonin Neurons in the Balance between Reward and Aversion. *International Journal of Molecular Sciences*, 21(6), 2160.
- Nawijn, L., van Zuiden, M., Koch, S. B. J., Frijling, J. L., Veltman, D. J., & Olff, M. (2016). Intranasal oxytocin enhances neural processing of monetary reward and loss in post-traumatic stress disorder and traumatized controls. *Psychoneuroendocrinology*, 66, 228 – 237.
- Newman, G. E., & Jeremy Shen, Y. (2012). The counterintuitive effects of thank-you gifts on charitable giving. *Journal of Economic Psychology*, 33(5), 973 – 983.
- Nosenzo, D., Quercia, S., & Sefton, M. (2015). Cooperation in small groups: The effect of group size. *Experimental Economics*, 18, 4 – 14.
- Noussair, C. N., van Soest, D., & Stoop, J. (2015). Punishment, reward, and cooperation in a framed field experiment. *Social Choice and Welfare*, 45(3), 537 – 559.
- Nunney, S. J., Schalk, J., & Manstead, A. S. R. (2022). Emotion and intergroup cooperation: How verbal expressions of guilt, shame, and pride influence behavior in a social dilemma. *Journal of Behavioral Decision Making*, 35(4), e2273.
- Orsini, C. A., Trotta, R. T., Bizon, J. L., & Setlow, B. (2015). Dissociable Roles for the Basolateral Amygdala and Orbitofrontal Cortex in Decision-Making under Risk of Punishment. *Journal of Neuroscience*, 35(4), 1368 – 1379.
- Ozono, H., Kamijo, Y., & Shimizu, K. (2017). Punishing second-order free riders before first-order free riders: The effect of pool punishment priority on cooperation. *Scientific Reports*, 7, 14379.
- Paula, J. R., Messias, J. P., Grutter, A. S., Bshary, R., & Soares, M. C. (2015). The role of serotonin in the modulation of cooperative behavior. *Behavioral Ecology*, 26(4), 1005 – 1012.
- Pedersen, E. J., McAuliffe, W. H. B., Shah, Y., Tanaka, H., Ohtsubo, Y., & McCullough, M. E. (2020). When and Why Do Third Parties Punish Outside of the Lab? A Cross-Cultural Recall Study. *Social Psychological and Personality Science*, 11(6), 846 – 853.
- Pei, H., Yan, G., & Wang, H. (2021). Reciprocal rewards promote the evolution of cooperation in spatial prisoner's dilemma game. *Physics Letters A*, 390, 127108.
- Pimentel, A. F. N., Carvalho, T. dos S., Lima, F., Lima – Maximino, M., Soares, M. C., & Maximino, C. (2019). Conditional approach as cooperation in predator inspection: A role for serotonin? *Behavioural Brain Research*, 365, 164 – 169.
- Rai, A., Ghosh, P., Chauhan, R., & Singh, R. (2018). Improving in-role and extra-role performances with rewards and recognition: Does engagement mediate the process? *Management Research Review*, 41(8), 902 – 919.
- Raihani, N. J., & Bshary, R. (2019). Punishment: One tool, many uses. *Evolutionary Human Sciences*, 1, 1 – 26.
- Rand, D. G., Armao IV, J. J., Nakamaru, M., & Ohtsuki, H. (2010). Anti-social punishment can prevent the co-evolution of punishment and cooperation. *Journal of Theoretical Biology*, 265(4), 624 – 632.
- Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive Interactions Promote Public Cooperation. *Science*, 325, 1272 – 1275.
- Rankin, D. J., Bargum, K., & Kokko, H. (2007). The tragedy of the commons in evolutionary biology. *Trends in Ecology & Evolution*, 22(12), 643 – 651.
- Rolls, E. T. (2019). The cingulate cortex and limbic systems for emotion, action, and memory. *Brain Structure and Function*, 224(9), 3001 – 3018.
- Salazar, M., Joel Shaw, D., Czekóová, K., Staněk, R., & Brázdil, M. (2022). The role of generalised reciprocity and reciprocal tendencies in the emergence of cooperative group norms. *Journal of Economic Psychology*, 90, 102520.
- Sefton, M., Shupp, R., & Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry*, 45(4), 671 – 690.
- Shiraki, Y., & Igarashi, T. (2018). “Paying it forward” via satisfying a basic human need: The need for relatedness satisfaction mediates gratitude and prosocial behavior. *Asian Journal of Social Psychology*, 21(1–2), 107 – 113.
- Stuewig, J., Tangney, J. P., Kendall, S., Folk, J. B., Meyer, C. R., & Dearing, R. L. (2015). Children's Proneness to Shame and Guilt Predict Risky and Illegal Behaviors in Young Adulthood. *Child Psychiatry & Human Development*, 46(2), 217 – 227.
- Strimling, P., & Eriksson, K. (2014). Regulating the regulation: Norms about punishment. In P. A. M. Van Lange, B. Rockenbach, & T. Yamagishi (Eds.), *Reward and punishment in social dilemmas* (pp. 52 – 69). New York, NY: Oxford.
- Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral Emotions and Moral Behavior. *Annual Review of Psychology*, 58, 345 – 372.
- Ule, A., Schram, A., Riedl, A., & Cason, T. N. (2009). Indirect Punishment and Generosity Toward Strangers. *Science*, 326, 1701 – 1704.
- van den Berg, P., Dewitte, P., Aertgeerts, I., & Wenseleers, T. (2020). How the incentive to contribute affects contributions in the one-shot public goods game. *Scientific Reports*, 10, 18732.
- Wang, L., & Murnighan, J. K. (2017). The dynamics of punish-

- ment and trust. *Journal of Applied Psychology*, 102(10), 1385 – 1402.
- Wang, S., Chen, X., & Szolnoki, A. (2019). Exploring optimal institutional incentives for public cooperation. *Communications in Nonlinear Science and Numerical Simulation*, 79, 104914.
- Wheelan, S. A. (2009). Group Size, Group Development, and Group Productivity. *Small Group Research*, 40(2), 247 – 262.
- Wu, J., Balliet, D., Peperkoorn, L. S., Romano, A., & Van Lange, P. A. M. (2020). Cooperation in Groups of Different Sizes: The Effects of Punishment and Reputation – Based Partner Choice. *Frontiers in Psychology*, 10, 2956.
- Wu, J., Luan, S., & Raihani, N. (2022). Reward, punishment, and prosocial behavior: Recent developments and implications. *Current Opinion in Psychology*, 44, 117 – 123.
- Wu, J. J., Zhang, B. Y., Zhou, Z. X., He, Q. Q., Zheng, X. D., Cressman, R., & Tao, Y. (2009). Costly punishment does not always increase cooperation. Proceedings of the National Academy of Sciences of the United States of America. punishment does not always increase cooperation. *Proceedings of the National Academy of Sciences*, 106(41), 17448 – 17451.
- Wu, Y., Chang, S., Zhang, Z., & Deng, Z. (2017). Impact of Social Reward on the Evolution of the Cooperation Behavior in Complex Networks. *Scientific Reports*, 7, 41076.
- Wu, Y., Zhang, Z., & Chang, S. (2019). Reciprocal reward promotes the evolution of cooperation in structured populations. *Chaos, Solitons & Fractals*, 119, 230 – 236.
- Yang, C. – L., Zhang, B., Charness, G., Li, C., & Lien, J. W. (2018). Endogenous rewards promote cooperation. *Proceedings of the National Academy of Sciences*, 115(40), 9968 – 9973.
- Ye, H., Chen, S., Luo, J., Tan, F., Jia, Y., & Chen, Y. (2016). Increasing returns to scale: The solution to the second – order social dilemma. *Scientific Reports*, 6, 31927.
- Yu, L., Chen, C., Niu, Z., Gao, Y., Yang, H., & Xue, Z. (2021). Risk aversion, cooperative membership and the adoption of green control techniques: Evidence from China. *Journal of Cleaner Production*, 279, 123288.
- Zelmer, J. (2003). Linear Public Goods Experiments: A Meta – Analysis. *Experimental Economics*, 6(3), 299 – 310.
- Zhang, B., An, X., & Dong, Y. (2021). Conditional cooperator enhances institutional punishment in public goods game. *Applied Mathematics and Computation*, 390, 125600.
- Zhang, Z., Song, F., & Song, Z. (2020). Promoting knowledge sharing in the workplace: Punishment v. reward. *Chaos, Solitons & Fractals*, 131, 109518.

The Patterns and Differences of the Effects of Rewards and Punishments in Social Dilemma Cooperation

Zhou Jingkun Cui Liying

(School of Psychology, Shanghai Normal University, Shanghai 200234)

Abstract: As an important incentive mechanism, although rewarding cooperators and punishing free riders both can promote cooperation in social dilemmas, they have difference in the modes and processes on cooperation. On the one hand, the effect of reward and punishment mechanism on cooperation is different, and the optimal effect of reward and punishment mechanism is affected by different reward and punishment environment. On the other hand, there are process differences in the influence of reward and punishment mechanism on cooperation. Reward and punishment mechanism can promote cooperation by inducing reciprocal mechanisms, emotional mechanisms, behavioral mechanisms, physiological mechanisms. At the same time, the reward and punishment mechanism will also destroy cooperation by inducing resource loss, cognitive loss and behavioral loss. In the future, the research can expand the impact of rewards and punishments on cooperative behavior in more types or situations, and increase the ecological validity of research conclusions, clarify the optimal intensity and group size of rewards and punishments on cooperation, and effectively identify and determine the boundaries between reward and punishment mechanisms in promoting and undermining cooperation.

Key words: reward; punishment; cooperation; social dilemma; boundaries