PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2013, 40(9): 796~803
www.pibb.ac.cn

人类的利他性惩罚：认知神经科学的视角 *

张耀华 [1, 2] 林珠梅 [1, 2] 朱莉琪 [1]**
([1] 中国科学院心理研究所行为科学重点实验室，北京 100101；[2] 中国科学院大学，北京 100049)

摘要 利他性惩罚广泛存在于人类社会中，在群体合作与规范维护方面起着重要的积极作用．个体作为潜在的惩罚者，从知觉到不公平事件到做出惩罚行为，需要经过一系列的认知和情绪过程，包括公平判断、奖赏加工、自我控制以及心理化等过程，并且调用相应的神经生理机制．认知神经科学为理解人类的利他性惩罚行为提供了新的视角和方法．本文基于最新的研究发现，综述了利他性惩罚相关的神经生理基础．

关键词 利他性惩罚，最后通牒博弈，第三方惩罚，功能性磁共振成像
学科分类号 B845.5，Q427 **DOI**: 10.3724/SP.J.1206.2012.00627

人类的合作行为长久以来受到研究者的广泛关注，然而既有的理论在解释上往往存在缺口，无法覆盖人类合作与利他的多样性和复杂性[1]．例如，即使人们不可能从帮助他人中获得好的声望，为什么仍然会帮助与其毫无血缘关系，且不会有进一步交往的陌生人？又如，人类社会为什么会存在如此大规模的合作行为，而其他物种却难以企及？最近，研究者把视角转向惩罚，开始探讨惩罚与合作之间的关系[2-3]，采用博弈游戏的范式观察个体的惩罚行为以及惩罚对各方收益与群体合作的影响[4-5]，并且提出利他性惩罚(altruistic punishment)这一概念来描述和概括相关的研究发现和理论构建．近年来，对利他性惩罚的研究已然成为心理学、社会学、经济学、进化生物学等众多学科的研究热点之一，随着神经科学的发现，研究者开始借助脑成像技术对利他性惩罚进行探讨[6-8]．

1 利他性惩罚的含义

利他性惩罚最初由瑞士苏黎世大学的 Fehr 教授等提出[4, 9]．在一系列实验中，Fehr 及其同事发现，人们倾向于惩罚不合作者(即分享公共资源，却不讲贡献的人)，而随着时间的推进，那些原本的不合作者也倾向于做出更多的贡献．他们采用的范式为公共品博弈(public goods game)，实验参加者在每一轮次开始时匿名决定自己的贡献额，其后研究者将每一名参加者的贡献额(公共品)汇总增值，再平均分配给参加者，而不管他们各自的贡献多少．以往的研究发现，随着时间的推进，实验参加者越来越倾向于少贡献，甚至不贡献，最终公共品耗竭，合作无以为继．然而，当研究者[4-5]在游戏中加入惩罚环节后，实验参加者在得知其他人员的贡献值之后，可以用手中的点币(token)减少他人的收益，他们发现贡献较多的个体(合作者)往往会使用点币用来减少贡献较少个体(搭便车者)的收益．也就是说，在博弈游戏中当惩罚成为可能时，对不合作者的惩罚普遍存在，进而惩罚促进了群体的合作．Fehr 等[4-5]在研究中把公共品博弈设计成单次(single shot)完全匿名，即在每一回合中，实验的参与者完全随机，且每一轮的搭档也完全不同．在这种情形下，个体并不能从有偿的惩罚中获得直接的收益，然而却要承担惩罚的成本，从功能的角度出发，有偿惩罚促进了群体的合作给他人带来了收益，在这个意义上，Fehr 及其同事将其称为利他性惩罚[4]．

* 国家重点基础研究发展计划(973)(2010CB8339004)、国家自然科学基金(30970911)和中科院重点部署项目(KJZD-EW-L04)资助．
** 通讯联系人．
Tel: 010-64836643, E-mail: zhulq@psych.ac.cn
收稿日期：2012-12-27，接受日期：2013-05-15

根据潜在惩罚者的收益是否相关可分为第二方惩罚和第三方惩罚[9]. 惩罚者作为利益攸关者所实施的惩罚为第二方惩罚，前述的公共品博弈中，由于不合作者的贡献行为影响到潜在惩罚者的收益，可被视为第二方惩罚；而在第三方惩罚中，惩罚者作为利益无关者实施惩罚行为. 有偿惩罚可以是直接或间接互惠的一种形式[10]. 然而，惩罚并非总是出于互惠.

对利他性惩罚的研究主要采用博弈游戏的范式. 在第二方惩罚的研究范式中，除了上面提到的公共品博弈，更多采用最后通牒博弈[11]. 在这个游戏中，玩家一作为提议者决定资源的分配(比如，如何分配 10 块钱)，玩家二作为回应者可以接受玩家一的分配方案，也可以拒绝. 如果回应者接受，资源按照分配方案进行；而一旦拒绝，则双方的收益均为零. 众多的研究表明，回应者并不总是接受所有的提议，而是拒绝低于总额 20%的出价[12]，表现为惩罚行为. 第二方惩罚还可采用囚徒困境博弈[10]的范式，其中在游戏设置上，如同公共品博弈一样，实验者会给予参试者一定的惩罚点币，用于惩罚不合作或背叛者.

第三方惩罚也有不同的研究范式[13-14]. 与第二方惩罚有所不同的是惩罚者作为利益无关方出现在游戏中，在目睹背叛或者破坏社会规范的事件后，潜在的惩罚者可以利用手中的惩罚点币减少过错方的收益，表现出第三方惩罚行为. 就具体的研究范式来说，包括独裁者博弈版[14]、囚徒困境版[14-15]、礼物赠予博弈版[16]以及信任博弈版[15, 17]的第三方惩罚. 以独裁者博弈为例，在典型的独裁者博弈中，“独裁者”一方可以任意决定资源如何分配(比如，如何分配 10 块钱)，接收方只能够默认所提出的分配方案，而没有拒绝的权力. 独裁者博弈版的第三方惩罚中，作为旁观者的第三方加入游戏中，目睹游戏中“独裁者”的分配，当给予接收方的资源严重少于“独裁者”所占有的资源时，第三方会利用手中的点币惩罚“独裁者”. 以这些范式进行研究发现，个体倾向于降低自己的收益来惩罚不公平分配者和不合作者. 最近，研究者以更灵活的方式考察第三方惩罚现象. Ohtsubo 等[18]发现，在博弈游戏中，作为第三方的参试者在观察到玩家发出欺骗性的信息时也倾向于对其施以惩罚.

在单次博弈中，纯粹理性的个体不应该做出利他性惩罚. 以最后通牒博弈为例，即使提议者给回应者的分配少于 20%，但是却要大于拒绝所带来的零收益，回应者从经济的角度考虑不应该加以拒绝. 然而，既有的研究和跨文化发现却指出利他性惩罚广泛存在于人类社会中[9, 19]. 利他性惩罚对于社会的合作以及社会规范的维护具有显要影响[20]. 尽管以往的实验室和实地研究加深了我们对利他性惩罚的理解，然而在对其进行理论解释上却存有争议. 研究者在解释利他性惩罚时常常聚焦于潜在惩罚者的不公平规避[21-22]或平等主义动机(egalitarian motive)[23]、负性情绪(愤怒)[4]、声望顾虑[24-25]以及提议者或者过错方的意图[26]等多种原因. 神经科学视角的加入有可能在神经水平上整合并扩展人们对利他性惩罚的认识.

2　利他性惩罚的神经基础

利他性惩罚涉及众多的脑区，而同样的脑区在不同的情形下往往又会具有不同的功能. 在这一部分，根据研究中的任务和条件设置，我们把利他性惩罚所涉及的心理成分划分为四个过程，即：公平判断、奖赏加工、自我控制以及心理化过程，并以这四个心理过程来组织利他性惩罚的神经基础. 然而，这些心理过程所涉及的神经基础却并非彼此独立，内外部刺激可以通过改变某些脑区以及脑区之间的连结来达到对利他性惩罚行为的影响(图 1). 接下来的部分将详述利他性惩罚所涉及的脑区，以及这些脑区在其中所发挥的功能.

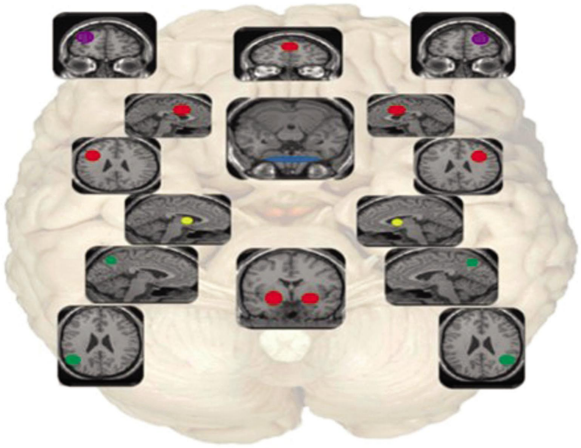

Fig. 1　The neural substrates of altruistic punishment

图 1　利他性惩罚主要脑区示意图

红色标注的脑区参与公平判断，自上而下分别为内侧前额皮层、扣带皮层、脑岛以及杏仁核；紫色标注的部分参与自我调节，主要为前额皮层；黄色标注的脑区参与奖赏加工，为伏隔核与尾状核所构成的纹状体；绿色标注的脑区参与心理化过程，涉及楔前叶和颞顶结合部. 利他性惩罚所涉及众多脑区并非彼此独立，而是相互作用调节着最后的惩罚行为[27-28]，而以蓝色标注的眶额皮层可能在最后的惩罚行为决策中起着调节作用[27, 29].

2.1 利他性惩罚需要公平判断

在对利他性惩罚的研究中，多数研究主要基于分配领域内的利他性惩罚，因而公平规范评价，以及与公平规范违背之后的负性情绪状态首先受到研究者的注意．其中涉及前脑岛(anterior insula，AI)、扣带皮层(cingulate cortex)以及与其相邻的内侧前额皮层(medial prefrontal cortex，mPFC)和边缘系统的杏仁核(amygdala)等．

前脑岛涉及对不公平(inequity)的加工[30]．然而，对于前脑岛在不公平加工中所扮演的具体角色还存在争议．一些研究者认为，前脑岛表征当前的情感状态，尤其是愤怒和厌恶情绪[31]．在面对不公平出价时，个体脑岛的激活反映了个体的负性情绪状态．既有的研究发现，前脑岛对出价不公平的程度敏感，随着不公平程度的增加，该脑区的激活程度也随之增强[32]．左侧前脑岛对不公平敏感，而不管不公平所指向的对象是自己，还是他人[33-34]．前脑岛激活程度越高的个体拒绝不公平出价的比例越大[32, 35]．右侧前脑岛对于随后被拒绝不公平出价的激活程度更高[32]，而如果随后的不公平出价被接受，则左侧前脑岛活动性往往会降低[35]．与负性情绪在前脑岛的表征相一致，不公平出价并不能增强老年人前脑岛的活动性[36]．

然而，另有研究者提出前脑岛编码更为一般的规范误差信号[37]．前脑岛的反应编码人际互动中社会规范的破坏．因而，脑岛的活动在某种程度上依赖于规范所在的背景，Güroğlu 等[38]发现，脑岛并不总是与拒绝不公平出价相关，当提议者所给出的分配方案违背了背景所暗含的规范时，接受该分配方案让脑岛的活动性增强．最近一项以佛教徒禅修者为实验对象的研究发现，相比于常人，禅修者在面对不公平出价时的前脑岛活动性并无异于公平出价[39]，从而表明前脑岛编码规范的违背，而非情绪的诱发．此外，脑岛后端(posterior insula)似乎与绝对的不公平有关，而脑岛中部(bilateral mid-insula)却起着整合不公平与背景的作用，因而与相对不公平有关[40]．

负性情绪状态同样激活了前扣带皮层(anterior cingulate cortex，ACC)以及内侧前额皮层(medial prefrontal cortex，mPFC)[41]．在 Sanfey，Rilling 及其同事的文章里[32]，背侧前扣带回也对不公平出价作出反应．随着提议者的出价偏离个体的期望值越来越大时，前扣带回的激活程度也越来越强[42]．与前述的前脑岛有所不同，扣带皮层似乎只对于指向自己的不公平敏感，包括中扣带皮层前端(anterior middle cingulate cortex，aMCC)[34]和双外侧扣带回(cingulate gyrus)[43]，与前扣带皮层相邻的另一脑区内侧前额皮层(mPFC)表现出类似的功能[33]，且对指向自己的劣势不公平有着更强的反应[34]．以往的研究者[44]指出，随着腹内侧前额皮层向前端延伸，其功能也越来越复杂，情绪加工以及与自我有关的过程都涉及该脑区．因而在最后通牒博弈中，该脑区有可能负责对不公平出价所诱发情绪的编码和调节．

此外，Gospic 等[45]发现，镇静类药物(benzodiazepine oxazepam)可以抑制杏仁核对不公平出价的活动，进而降低个体的利他性惩罚行为，表现为对不公平出价的拒绝率下降．对于杏仁核损伤患者来说，虽然他们的公平判断并无异常，但却表现出全或无的反应模式，即对于不公平出价的全盘接受以及对于极端不公平高出价的全盘拒绝[46]．

Buckholtz 等[47]在有关司法领域内第三方惩罚的脑机制研究中发现，对被告所施以的惩罚幅度同样由与社会和情感加工的脑区相关，包括右侧杏仁核、后扣带皮层以及内侧前额皮层．

2.2 利他性惩罚需要奖赏加工

从功能的角度，利他性惩罚促进群体合作与规范维护．不过从主观角度，利他性惩罚却可以激活与奖赏加工相关的脑区．Singer 等[48]发现，对背叛者施加疼痛刺激不仅让男性受害者有关疼痛共情加工的脑区激活程度变弱，同时还激活了左侧腹侧纹状体的伏隔核(nucleus accumbens，NAcc)以及左侧眶额皮层(orbitofrontal cortex，mOFC)变得更为活跃．有关利他性惩罚激活奖赏中心更为直接的证据来自 Fehr 教授研究小组的发现[49]．通常，被信任者的互惠行为被视为一种社会规范，对这种规范的破坏往往招致惩罚[50]．在实验中，研究者赋予作为信任者的个体可以行使惩罚被信任者背叛(不互惠)行为的权利，在得知被信任者的背叛行为后，当背叛者的背叛意图清晰或者惩罚有效时，个体所作出的惩罚激活了尾状核(nucleus caudatus, NCd)[49]．该脑区作为背侧纹状体(dorsal striatum, dSTR)的一组成部分编码着目标指向的奖赏行为，因而与对奖赏的预期有关[51]，预期施加惩罚能够带来满意感让个体作出了更大的惩罚行为，即使惩罚需要付出代价．

与第二方惩罚激活奖赏中枢类似，在第三方惩罚中也有同样的发现．Baumgartner 等[27]报道，在对外群体成员的惩罚中，右侧尾状核的活动性增

强．Civai 等[34]发现，奖赏中枢的激活还受到其他变量的调节，无论是在第二方还是第三方惩罚中，虽然惩罚都激活了尾状核和伏隔核．但是尾状核对于有效惩罚保持着高度敏感，而伏隔核的激活尤其表现在个体自己遭受不公平时所作出的惩罚．

利他性惩罚者从惩罚背叛者或者不公平者中获得满足感，即所谓“甜蜜的复仇”．这一发现也被视为利他性惩罚得以进化的近端机制．然而，最近的行为研究表明，虽然人们预期从惩罚中获得享乐奖赏，而实际上复仇或者惩罚随后让人们感觉更糟[52]，享乐主义情感调节(hedonic affect regulation)并非驱动人们作出利他性惩罚行为的心理机制[53]．如何调和上述发现，以及获得更多有关情感动力学变化的知识需要未来更多的研究．

2.3 利他性惩罚需要自我控制

前额皮层通常卷入与自我调节和情绪调节有关的活动中，因而该脑区可以通过调节参与公平判断的脑区以及参与奖赏和价值加工的脑区来改变最终的利他性惩罚行为．Sanfey，Rilling 及其同事发现[32]，相比于公平出价，不公平出价激活了背外侧前额皮层(dorsal lateral prefrontal cortex，DLPFC)，但是该脑区对不公平出价维持着相对的稳定性，它的激活并不能预测个体的拒绝率．Wright 等[40]发现，相比于不公平背景下的分配，拒绝公平背景下的同一出价让右侧 DLPFC 的激活程度更高．尤其当惩罚有效性很弱(惩罚成本较高)时，个体更需要调用右侧 DLPFC 来完成惩罚[43]．Buckholtz 等[47]通过改变罪行过失的严重性让参试者作出惩罚判决，以检验司法领域内第三方惩罚的脑机制．他们发现，右侧 DLPFC 的活动性与惩罚判决有关，惩罚判决让该脑区的活动性增强，而不管当事人是否对所犯罪行负有责任．

反过来，研究者发现抑制右侧前额皮层可以降低个体的利他性惩罚行为，即，对不公平出价表现出较高的接受率[54]．结合重复性经颅磁刺激(repetitive transcranial magnetic stimulation，rTMS)和脑成像方法的研究表明，右脑背外侧前额皮层与腹内侧前额皮层后端(posterior ventromedial prefrontal cortex，pVMPFC)之间存在交互[55]．以往的研究发现，腹内侧前额皮层后端与决策价值有关[56]，因而在面对不公平出价时，自利和公平动机之间的冲突可能需要这两个脑区的互动，并且通过后者编码拒绝还是接受的决策价值(decision value)进而作出利他性惩罚．在第三方利他性惩罚中，前额皮层对价值加工脑区的调节还受到群体边界的影响．第三方惩罚中右外侧前额皮层还可以增强其与右侧眶额回(right orbitofrontal gyrus)的联系，决定对外群体的背叛者作出惩罚[27]．

额叶的其他脑区同样参与了利他性惩罚中．尽管在 Sanfey 等[32]的研究中，背外侧前额皮层与个体的拒绝率之间不存在关系，但是 Güroğlu 等[38]却发现，在第二方惩罚中，拒绝相比于接受不公平出价显著地激活了左侧 DLPFC，并且还不受提议者备选项变量的调节．尽管如此，对左侧 DLPFC 的调用还受到不公平所指对象的影响，当个体作为受害者，如果他选择不作出惩罚，需要更多地调用左侧 DLPFC，而当个体作为旁观者，如果他选择进行惩罚，左侧 DLPFC 的活动性增强[43]．同时，对左侧 DLPFC 的调用还存在年龄差异，尽管在年轻人中不存在左侧 DLPFC 激活与对不公平出价的接受率之间的关系，老年人在这一脑区的激活程度显著预测了他们对极端不公平出价的接受比率[36]．

类似地，相比于拒绝或者惩罚不公平出价对背侧前额皮层活动性的效应，Tabibnia 等[35]研究显示，接受不公平出价需要右脑腹外侧前额皮层(ventrolateral prefrontal cortex，VLPFC)，并且通过减弱左侧脑岛的活动性，来提高个体不公平出价的接受比例，这一点与背侧前额皮层活动幅度无法预测拒绝率不同．另一个与自我调节有关的脑区为眶额皮层(orbitofrontal cortex，OFC)[57]，Mehta 与 Beer 发现[58]，内侧眶额皮层的激活程度越高，个体越倾向于接受不公平出价，活动水平越低，个体越倾向于作出惩罚拒绝．

前额皮层的不同脑区似乎涉及自我调节和执行控制的不同方面，表现为左右脑区功能的不对称性[59]、各子区功能的特异性[60]．在利他性惩罚中，前额皮层的不同脑区也表现出类似的功能划分．究竟沿着哪条轴线来组织前额皮层在认知和行为控制上的功能尚存在争议[61]，鉴于前额皮层在利他性惩罚行为表现上的复杂性，因而需要更多的研究来深化我们对前额皮层在社会决策，尤其利他性惩罚上的认识．

2.4 利他性惩罚需要心理化过程

心理化(mentalizing)这一概念最初来自比较和发展心理学，指的是个体对他人心理状态的理解与操纵，进而达到解释、预测和影响他人行为的目的[62]，也有研究者以心理理论(theory of mind, ToM)来代指个体所获得的这种能力[63]，它表明个体开始

从信念和意图等心理状态来理解他人的行为．相关的行为研究表明，提议者出价的心理状态，即意图影响着人们对出价的反应．当公平规范突显时，即备选项为公平分配，个体意图的效应最为明显[64]．例如，当提议者面对两个选项，在一种情形下，其中一个选项为公平分配，另一选项为不公平分配，而在另一种情形下，其中一个选项为不公平分配，另一选项为更不公平分配．虽然提议者在两种情形下都选择了不公平分配，但是人们倾向于对第一种情形下的不公平分配作出更大的惩罚．

Rilling，Sanfey 及其同事[28]采用最后通牒博弈和囚徒困境博弈范式，通过对比控制条件(保持互动性质，但对家为计算机，或者取消对家，仅保留收益结果)，研究者发现，在每轮游戏收益呈现时，除了经典的与心理化有关的脑区，如，副扣带皮层前端(anterior paracingulate cortex)、右侧颞上沟后端(posterior superior temporal sulcus，pSTS)得到激活之外，楔前叶(precuneus)、颞上沟中部(mid STS)以及海马(hippocampus)的激活程度也显著提高．当提议者的选择所暗含的意图不明显时，拒绝不公平出价需要调用与心理理论有关的脑区，反映在神经水平上，当回应者拒绝不公平出价时，右侧颞顶结合部(temporoparietal junction，TPJ)、内侧前额皮层以及前扣带回的激活程度显著提高[38]．而颞顶结合部的激活程度还可以解释为什么随着年龄的增长个体倾向于接受来自于提议者无法选择的不公平出价[65]．Wright 等[40]发现，个体对公平的比较成分(不公平差距)越关注，就越倾向于在面对出价的时候调用与心理理论有关的脑区，包括楔前叶和颞顶结合部等．

在司法领域内的第三方惩罚中，如果有可减轻罪行的情节，比如，被告人在遭遇家暴、疾病或者贫穷等情况下所犯下的罪行往往会得到人们的同情，反映在脑活动上背内侧前额皮层(dorsomedial prefrontal cortex，DMPFC)、楔前叶以及颞顶结合部[47]等脑区的活动程度显著增加，而减刑判决则激活了前扣带回，并且与个体右侧脑岛中部的活动性有关[66]．Baumgartner 等[27]发现，个体为了减少对内群体背叛成员的惩罚，往往激活与心理化过程有关的脑区，包括背内侧前额皮层以及双外侧颞顶结合部，而且随着惩罚力度的减弱，这两个脑区之间的连接强度则趋于加强．进一步的分析显示，左侧颞顶结合部调节着背内侧前额皮层与惩罚相关脑区的关系，包括右侧眶额回等与价值加工相关的脑区，从而让个体减少对内群体成员的惩罚．

利他性惩罚涉及众多脑区，而同样的脑区在不同的情形下又会起到不同的功能．在这一部分，根据研究中的任务和条件设置，我们把利他性惩罚划分为四个心理过程．然而，这些心理过程所涉及的神经基础却并非彼此独立，内外部刺激可以通过改变某些脑区以及脑区之间的连结来达到对利他性惩罚行为的影响．

3 小结与展望

个体作出利他性惩罚看似简单，实际上涉及多重心理和神经过程，以及相应的神经化学变化．利他性惩罚决策受到人格与社会情境因素的影响．尽管在过去 30 年间，尤其是近 10 年对利他性惩罚的研究已经取得显著的进步，然而仍有问题摆在研究者的面前．

首先，有关利他性惩罚脑机制的研究已向我们显示，众多的脑区卷入其中，然而各个脑区之间的联系以及如何交互作用决定最后的惩罚决策还知之甚少．最近，Kim 等[29]在研究提议者面孔可信任性对惩罚行为的影响中发现，同样的出价如果来自高面孔可信任的提议者，其被拒绝的可能性更低．在神经水平上，外侧眶额皮层(lateral orbitofrontal cortex，lOFC)与该决策偏差有关，在拒绝同样出价时，外侧眶额皮层的活动性增强，其与右脑杏仁核后端、左侧前脑岛、背侧前扣带回以及右脑背外侧前额皮层等的沟通也在加大，可见脑区之间的交互决定着最后的惩罚行为．与最近对社会行为的遗传学研究[67]取向相一致，研究者已经开始从双生子[68]和基因受体[43, 69]的角度开始对利他性惩罚进行研究．随着技术手段的革新，我们对利他性惩罚的了解将更为全面和立体．

第二，惩罚并非总是出于利他动机．虽然从惩罚者承担惩罚成本，且促进群体合作的意义上，有偿惩罚可以称为利他性惩罚[4]．然而，惩罚者做出惩罚行为其潜在的心理动机却异常复杂．即使惩罚指向背叛者或不公平分配者，惩罚也并非简单地出于利他[70]，例如，公平品博弈范式下的惩罚者会隐藏他们对背叛者的严厉惩罚[71]，因而表明惩罚可能出于自利、恶意或者竞争性动机[72-73]．另一方面，与利他性惩罚背道而行的另一惩罚为反社会性惩罚[74]，即损人不利己行为．反社会性惩罚的对象往往指向合作者或者利他性惩罚者，它的出现会妨碍惩罚与合作的进化[75]．更全面地理解有偿惩罚需

要了解惩罚背后的心理动机及其反社会性惩罚的神经生物机制，从而在神经水平上分离出惩罚的不同机制.

第三，惩罚并非总是促进群体合作. 惩罚对合作的积极作用受到惩罚成本[76]以及文化背景[77]的影响. 相比于单纯的惩罚，结合奖赏的惩罚更能够促进提议者作出公平出价[78]. 另一方面，相比于惩罚背叛者，个体还可以选择补偿受害者[79]，然而这方面的研究还处于起步阶段. 有关于惩罚背叛者、奖赏合作者以及补偿受害者的影响因素及其神经机制需要更多的研究.

第四，利他性惩罚作为一种重要的社会决策行为，涉及公平规范以及人际交互等诸多特征，因而可用以探讨其在异常人群中的表现，以便更好地了解社会行为缺陷的潜在机制[80]. 目前，已有研究者开始从这个角度研究精神分裂症患者[81]、精神病态患者[82]以及抑郁障碍患者[83]在利他性惩罚上表现. 反过来，在异常人群中的发现也可以厘清利他性惩罚的概念本身，例如，Wischniewski 与 Brüne[81]发现，精神分裂症患者在第三方惩罚上与常人无异，却表现出更少的第二方惩罚，即接受了更多的不公平出价，从而表明利他性惩罚两种类型的差异.

衍生于经济博弈的利他性惩罚范式虽然看似简单，却成为研究者深入探讨人类合作与规范履行的有力工具，对利他性惩罚行为的研究与发现必将加深我们对于人类社会生活的理解.

参 考 文 献

[1] Nowak M A. Five rules for the evolution of cooperation. Science, 2006, **314**(5805): 1560−1563

[2] Boyd R, Richerson P J. Punishment allows the evolution of cooperation (or anything else) in sizable groups. Ethology and Sociobiology, 1992, **13**(3): 171−195

[3] Gintis H. Strong reciprocity and human sociality. J Theoret Biol, 2000, **206**(2): 169−179

[4] Fehr E, Gächter S. Altruistic punishment in humans. Nature, 2002, **415**(6868): 137−140

[5] Fehr E, Gächter S. Cooperation and punishment in public goods experiments. Amer Econ Rev, 2000, **90**(4): 980−994

[6] Sanfey A G. Social decision-making: insights from game theory and neuroscience. Science, 2007, **318**(5850): 598−602

[7] Fehr E, Camerer C F. Social neuroeconomics: the neural circuitry of social preferences. Tren Cogn Sci, 2007, **11**(10): 419−427

[8] Seymour B, Singer T, Dolan R. The neurobiology of punishment. Nat Rev Neurosci, 2007, **8**(4): 300−311

[9] Fehr E, Fischbacher U. The nature of human altruism. Nature, 2003, **425**(6960): 785−791

[10] Dreber A, Rand D G, Fudenberg D, *et al*. Winners don't punish. Nature, 2008, **452**(7185): 348−351

[11] Güth W, Schmittberger R, Schwarze B. An experimental analysis of ultimatum bargaining. J Econom Behav Organ, 1982, **3**(4): 367−388

[12] Camerer C, Thaler R H. Ultimatums, dictators and manners. J Econom Perspectives, 1995, **9**(2): 209−219

[13] Kahneman D, Knetsch J L, Thaler R H. Fairness and the assumptions of economics. Journal of Business, 1986, **59** (04): S285−S300

[14] Fehr E, Fischbacher U. Third-party punishment and social norms. Evolution and Human Behavior, 2004, **25**(2): 63−87

[15] Kurzban R, Descioli P, O'brien E. Audience effects on moralistic punishment. Evolution and Human Behavior, 2007, **28**(2): 75−84

[16] Shinada M, Yamagishi T, Ohmura Y. False friends are worse than bitter enemies: "Altruistic" punishment of in-group members. Evolution and Human Behavior, 2004, **25**(6): 379−393

[17] Charness G, Cobo-Reyes R, Jiménez N. An investment game with third-party intervention. J Economic Behavior & Organization, 2008, **68**(1): 18−28

[18] Ohtsubo Y, Masuda F, Watanabe E, *et al*. Dishonesty invites costly third-party punishment. Evolution and Human Behavior, 2010, **31**(4): 259−264

[19] Henrich J, Mcelreath R, Barr A, *et al*. Costly punishment across human societies. Science, 2006, **312**(5781): 1767−1770

[20] Fehr E, Fischbacher U. Social norms and human cooperation. Trends in Cognitive Sciences, 2004, **8**(4): 185−190

[21] Loewenstein G F, Thompson L, Bazerman M H. Social utility and decision making in interpersonal contexts. J Pers Soc Psychol, 1989, **57**(3): 426−441

[22] Fehr E, Schmidt K M. A theory of fairness, competition, and cooperation. Quart J Econom, 1999, **114**(3): 817−868

[23] Fowler J H, Johnson T, Smirnov O. Human behaviour: egalitarian motive and altruistic punishment. Nature, 2005, **433**(7021): E1−E1

[24] Nowak M A, Page K M, Sigmund K. Fairness versus reason in the ultimatum game. Science, 2000, **289**(5485): 1773−1775

[25] Dos Santos M, Rankin D J, Wedekind C. The evolution of punishment through reputation. Proceedings of the Royal Society B: Biological Sciences, 2011, **278**(1704): 371−377

[26] Falk A, Fischbacher U. A theory of reciprocity. Games and Economic Behavior, 2006, **54**(2): 293−315

[27] Baumgartner T, Götte L, Gügler R, *et al*. The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. Human Brain Mapping, 2012, **33**(6): 1452−1469

[28] Rilling J K, Sanfey A G, Aronson J A, *et al*. The neural correlates of theory of mind within interpersonal interactions. NeuroImage, 2004, **22**(4): 1694−1703

[29] Kim H, Choi M-J, Jang I-J. Lateral OFC activity predicts decision bias due to first impressions during ultimatum Games. J Cogn Neurosci, 2011, **24**(2): 428−439

[30] Hsu M, Anen C, Quartz S R. The Right and the good: distributive justice and neural encoding of equity and efficiency. Science, 2008,

320(5879): 1092-1095

[31] Lamm C, Singer T. The role of anterior insular cortex in social emotions. Brain Structure and Function, 2010, **214**(5): 579-591

[32] Sanfey A G, Rilling J K, Aronson J A, *et al*. The neural basis of economic decision-making in the ultimatum game. Science, 2003, **300**(5626): 1755-1758

[33] Corradi-Dell'acqua C, Civai C, Rumiati R I, *et al*. Disentangling self- and fairness-related neural mechanisms involved in the ultimatum game: an fMRI study. Soc Cogn Affect Neurosci, 2013, **8**(4): 424-431

[34] Civai C, Crescentini C, Rustichini A, *et al*. Equality versus self-interest in the brain: Differential roles of anterior insula and medial prefrontal cortex. NeuroImage, 2012, **62**(1): 102-112

[35] Tabibnia G, Satpute A B, Lieberman M D. The sunny side of fairness. Psychological Science, 2008, **19**(4): 339-347

[36] Harlé K M, Sanfey A G. Social economic decision-making across the lifespan: An fMRI investigation. Neuropsychologia, 2012, **50**(7): 1416-1424

[37] Montague P R, Lohrenz T. To detect and correct: norm violations and their enforcement. Neuron, 2007, **56**(1): 14-18

[38] Güroğlu B, Van Den Bos W, Rombouts S a R B, *et al*. Unfair? It depends: neural correlates of fairness in social context. Soc Cogn Affect Neurosci, 2010, **5**(4): 414-423

[39] Kirk U, Downar J, Montague P R. Interoception drives increased rational decision-making in meditators playing the Ultimatum Game. Frontiers in Neuroscience, 2011, **5**(49): 1-11

[40] Wright N D, Symmonds M, Fleming S M, *et al*. Neural segregation of objective and contextual aspects of fairness. The Journal of Neuroscience, 2011, **31**(14): 5244-5252

[41] Etkin A, Egner T, Kalisch R. Emotional processing in anterior cingulate and medial prefrontal cortex. Trends in Cognitive Sciences, 2011, **15**(2): 85-93

[42] Chang L J, Sanfey A G. Great expectations: neural computations underlying the use of social norms in decision-making. Soc Cogn Affect Neurosci, 2013, **8**(3): 277-284

[43] Strobel A, Zimmermann J, Schmitz A, *et al*. Beyond revenge: neural and genetic bases of altruistic punishment. NeuroImage, 2011, **54**(1): 671-680

[44] Amodio D M, Frith C D. Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci, 2006, **7**(4): 268-277

[45] Gospic K, Mohlin E, Fransson P, *et al*. Limbic justice—amygdala involvement in immediate Rejection in the Ultimatum Game. PLoS Biol, 2011, **9**(5): e1001054

[46] Scheele D, Mihov Y, Kendrick K M, *et al*. Amygdala lesion profoundly alters altruistic punishment. Biological Psychiatry, 2012, **72**(3): e5-e7

[47] Buckholtz J W, Asplund C L, Dux P E, *et al*. The neural correlates of third-party punishment. Neuron, 2008, **60**(5): 930-940

[48] Singer T, Seymour B, O'doherty J P, *et al*. Empathic neural responses are modulated by the perceived fairness of others. Nature, 2006, **439**(7075): 466-469

[49] De Quervain D J F, Fischbacher U, Treyer V, *et al*. The neural basis of altruistic punishment. Science, 2004, **305**(5688): 1254-1258

[50] Bicchieri C, Xiao E, Muldoon R. Trustworthiness is a social norm, but trusting is not. Politics, Philosophy & Economics, 2011, **10**(2): 170-187

[51] O'doherty J, Dayan P, Schultz J, *et al*. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science, 2004, **304**(5669): 452-454

[52] Carlsmith K M, Wilson T D, Gilbert D T. The paradoxical consequences of revenge. J Personal Soc Psychol, 2008, **95**(6): 1316-1324

[53] Gollwitzer M, Bushman B J. Do victims of injustice punish to improve their mood?. Soc Psychol Personal Sci, 2012, **3**(5): 572-580

[54] Knoch D, Pascual-Leone A, Meyer K, *et al*. Diminishing reciprocal fairness by disrupting the right prefrontal cortex. Science, 2006, **314**(5800): 829-832

[55] Baumgartner T, Knoch D, Hotz P, *et al*. Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. Nature Neuroscience, 2011, **14**(11): 1468-1474

[56] Smith D V, Hayden B Y, Truong T-K, *et al*. Distinct value signals in anterior and posterior ventromedial prefrontal cortex. J Neurosci, 2010, **30**(7): 2490-2495

[57] Beer J S, Shimamura A P, Knight R T. Frontal Lobe Contributions to Executive Control of Cognitive and Social Behavior [M]// GAZZANIGA M S. The cognitive neurosciences (3rd ed). Cambridge, MA, US: MIT Press. 2004: 1091-1104

[58] Mehta P H, Beer J. Neural mechanisms of the testosterone-aggression relation: the role of orbitofrontal cortex. J Cogn Neurosci, 2009, **22**(10): 2357-2368

[59] Levy B J, Wagner A D. Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. Annals of the New York Academy of Sciences, 2011, **1224**(1): 40-62

[60] Chikazoe J. Localizing performance of go/no-go tasks to prefrontal cortical subregions. Curr Opin Psych, 2010, **23**(3): 267-272

[61] O'reilly R C. The what and How of prefrontal cortical organization. Trends in Neurosciences, 2010, **33**(8): 355-361

[62] Frith C D, Frith U. Interacting minds——a biological basis. Science, 1999, **286**(5445): 1692-1695

[63] Singer T. Understanding others: brain mechanisms of theory of mind and empathy [M]//GLIMCHER P W, CAMERER C F, FEHR E, *et al*. Neuroeconomics: Decision making and the brain. San Diego, CA, US: Elsevier Academic Press. 2009: 251-268

[64] Falk A, Fehr E, Fischbacher U. On the nature of fair behavior. Economic Inquiry, 2003, **41**(1): 20-26

[65] Güroğlu B, Van Den Bos W, Van Dijk E, *et al*. Dissociable brain networks involved in development of fairness considerations: Understanding intentionality behind unfairness. NeuroImage, 2011, **57**(2): 634-641

[66] Yamada M, Camerer C F, Fujie S, *et al*. Neural circuits in the brain that are activated when mitigating criminal sentences. Nature

Communications, 2012, **3**(759): 1−6

[67] Ebstein R P, Israel S, Chew S H, *et al*. Genetics of human social behavior. Neuron, 2010, **65**(6): 831−844

[68] Wallace B, Cesarini D, Lichtenstein P, *et al*. Heritability of ultimatum game responder behavior. Proc Natl Acad Sci USA, 2007, **104**(40): 15631−15634

[69] Zhong S, Israel S, Shalev I, *et al*. Dopamine D4 receptor gene associated with fairness preference in ultimatum Game. PLoS ONE, 2010, **5**(11): e13765

[70] Bshary R, Raihani N J. Toward an experimental exploration of the complexity of human social interactions. Proc Natl Acad Sci USA, 2011, **108**(45): 18195−18196

[71] Rockenbach B, Milinski M. To qualify as a social partner, humans hide severe punishment, although their observed cooperativeness is decisive. Proc Natl Acad Sci USA, 2011, **108**(45): 18307−18312

[72] Jensen K. Punishment and spite, the dark side of cooperation. Philosophical Transactions of the Royal Society B: Biological Sciences, 2010, **365**(1553): 2635−2650

[73] Masclet D, Noussair C, Tucker S, *et al*. Monetary and nonmonetary punishment in the voluntary contributions mechanism. Amer Econ Rev, 2003, **93**(1): 366−381

[74] Herrmann B, Thöni C, Gächter S. Antisocial punishment across societies. Science, 2008, **319**(5868): 1362−1367

[75] Rand D G, Armao Iv J J, Nakamaru M, *et al*. Anti-social punishment can prevent the co-evolution of punishment and cooperation. J Theor Biol, 2010, **265**(4): 624−632

[76] Egas M, Riedl A. The economics of altruistic punishment and the maintenance of cooperation. Proceedings of the Royal Society B: Biological Sciences, 2008, **275**(1637): 871−878

[77] Wu J-J, Zhang B-Y, Zhou Z-X, *et al*. Costly punishment does not always increase cooperation. Proc Natl Acad Sci USA, 2009, **106**(41): 17448−17451

[78] Andreoni J, Harbaugh W, Vesterlund L. The carrot or the stick: rewards, punishments, and cooperation. Amer Econ Rev, 2003, **93**(3): 893−902

[79] Leliveld M C, Van Dijk E, Van Beest I. Punishing and compensating others at your own expense: The role of empathic concern on reactions to distributive injustice. Euro J Soc Psychol, 2012, **42**(2): 135−140

[80] Kishida K T, King-Casas B, Montague P R. Neuroeconomic approaches to mental disorders. Neuron, 2010, **67**(4): 543−554

[81] Wischniewski J, Brüne M. Moral reasoning in schizophrenia: an explorative study into economic decision making. Cognitive Neuropsychiatry, 2011, **16**(4): 348−363

[82] Koenigs M, Kruepke M, Newman J P. Economic decision-making in psychopathy: a comparison with ventromedial prefrontal lesion patients. Neuropsychologia, 2010, **48**(7): 2198−2204

[83] Destoop M, Schrijvers D, De Grave C, *et al*. Better to give than to take? Interactive social decision-making in severe major depressive disorder. J Affect Disord, 2012, **137**(1−3): 98−105

The Human Altruistic Punishment: From The Perspective of Cognitive Neuroscience*

ZHANG Yao-Hua[1,2], LIN Zhu-Mei[1,2], ZHU Li-Qi[1)**]

([1)] *Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing* 100101, China;
[2)] *University of Chinese Academy of Sciences, Beijing* 100049, China)

Abstract Altruistic punishment is very common in human life, which is an effective mechanism enforcing group cooperation and social norm. From perceiving the transgressions to administering costly punishment, potential punishing requires a series of cognitive and affective processes, including fairness judgment, reward processing, self-control, mentalizing, *etc*., and their underlying neurophysiological mechanisms. Cognitive neuroscience provides alternative perspective and paradigm for understanding human punishment behavior. Based on the latest findings, the neurophysiological mechanisms of altruistic punishment are reviewed and discussed in this article.

Key words altruistic punishment, ultimatum game, third-party punishment, functional magnetic resonance imaging

DOI: 10.3724/SP.J.1206.2012.00627

* This work was supported by grants from National Basic Research Program of China (2010CB8339004), The National Natural Science Foundation of China (30970911) and the Key Research Program of the Chinese Academy of Sciences (KJZD-EW-L04).

**Corresponding author.

Tel: 86-10-64836643, E-mail: zhulq@psych.ac.cn

Received: December 27, 2012 Accepted: May 15, 2013