Advances in Psychology 心理学进展, 2023, 13(1), 32-36
Published Online January 2023 in Hans. https://www.hanspub.org/journal/ap
https://doi.org/10.12677/ap.2023.131005

Hans汉斯

第三方惩罚行为决策及其认知神经机制基础探析

敖丽红

华北理工大学心理与精神卫生学院，河北 唐山

收稿日期：2022年11月21日；录用日期：2023年1月5日；发布日期：2023年1月13日

摘　要

社会规范是人类社会长期发展和进步的基石，也是人类赖以生存的基本行为准则，社会规范的维护依赖于惩罚的执行。第三方惩罚是指与利益无关的第三方牺牲自己的利益对违背社会规范的人实施的惩罚行为，相较于其他惩罚行为，第三方惩罚更具客观、稳定和合理性。文章对第三方惩罚相关的部分研究进行了简要的梳理和总结，针对第三方惩罚的发生机制和神经基础进行了一定的阐述，以期为未来有关第三方惩罚的行为研究和脑机制研究提供一定的参考。

关键词

第三方惩罚，社会规范，利他惩罚，认知神经机制

Third-Party Punishment Behavioral Decision-Making and Its Cognitive Neural Mechanism Basic Analysis

Lihong Ao

School of Psychology and Mental Health, North China University of Science and Technology, Tangshan Hebei

Received: Nov. 21st, 2022; accepted: Jan. 5th, 2023; published: Jan. 13th, 2023

Abstract

Social norms are the cornerstone of the long-term development and progress of human society, and they are also the basic code of conduct for human survival, and the maintenance of social norms depends on the implementation of punishment. Third-party punishment refers to the punishment

文章引用: 敖丽红(2023). 第三方惩罚行为决策及其认知神经机制基础探析. *心理学进展, 13(1),* 32-36.
DOI: 10.12677/ap.2023.131005

of a person who violates social norms by a third party that has nothing to do with interests at the expense of his or her own interests, and third-party punishment is more objective, stable and reasonable than other punishments. This paper briefly sorts out and summarizes some researches related to third-party punishment, and elaborates on the occurrence mechanism, behavioral research and related neural mechanisms of third-party punishment, in order to provide some reference for future behavioral research and brain mechanism research on third-party punishment.

Keywords

Third-Party Punishment, Social Norms, Altruistic Punishment, Cognitive Neural Mechanism

1. 前言

人类社会在长期的进化和发展过程中，逐步形成了一些稳定而有效的社会规则，用以维护正常的社会生产和社会活动的进行。社会规范(Social Norms)，是指特定情境下某一群体成员广泛认可的行为标准(Fehr & Fischbacher, 2004a)。不论是国家还是个人，通常会通过惩罚违规者的方式去制止违反社会规范的事件的发生。因此，惩罚机制的存在对于维护社会准则、维系社会和谐稳定及发展皆具有重要意义。维护社会规范通常涉及三个主体，违规者，是违反社会规范或者侵犯受害者利益的行为主体；受害者，即因违规者的行为自身利益直接受损的主体，又称为第二方；第三方，即对违反规范的整个过程十分了解同时自身利益并未收到损害的旁观者。因此，根据做出维护社会规范的行为者所处的位置，将利他惩罚(Altruistic Punishment)分为了第二方惩罚和第三方惩罚。从维护社会规范的角度出发，第二方作为当事人，更容易受到个人因素的影响，因此，与之相比第三方的行为更加客观、稳定且合理，也更具有规范性(Bendor & Swistak, 2001)。

有关第三方惩罚的研究主要集中于经济学和心理学领域。只有在维护规范的第三方惩罚这一行为决策有足够深刻的认识的基础上，才能对人类社会的发展进行更深入的剖析。当然，人们会质疑在生活中第三方惩罚并不是经常出现，这是因为这种惩罚需要个体牺牲自我利益，与“经济人”理论并不完全相符合，该理论认为人们总是理性的并且追求利益最大化。然而，事实表明，每个人都有维护社会规范的意愿且绝大多数人都愿意付出成本去惩罚破坏社会规范的人。正如那句谚语——“路见不平、拔刀相助”，关键在于这种惩罚比较特殊，具有利他性。因此，以第三方为核心，探究其在维护社会规范中的行为决策以及背后的认知神经机制具有十分重要的现实意义。

2. 第三方惩罚行为决策及产生机制

维护社会规范可以由人际互动中的第二方或第三方来执行，第三方为维护社会规范所采取的惩罚被称为第三方惩罚，即当违规行为与自身利益无关时，个体牺牲自我利益来惩罚违规者的现象(Fehr & Fischbacher, 2004b)。现有研究主要聚焦于第二方惩罚，无论是惩罚动机还是后果，均已受到研究者的长期关注和调查。相较之下，作为利他惩罚中的一种，具有一定特殊性的第三方惩罚还需进行深入探索。研究者们经常采用一些经典的经济博弈任务对第三方惩罚进行研究，最常见的有独裁者博弈任务(Dictator Games, DG)、最后通牒博弈任务(Ultimatum Game, UG) (Güth et al., 1982)、囚徒困境任务(Prisoner’s Dilemma game, PD) (Fehr & Fischbacher, 2004b)或公共物品博弈任务(Public Good game, PG) (Fehr & Gächter, 2002)。

有研究者认为人们对违规者实施惩罚，是因为违规者的“罪有应得”，通过惩罚也能够直接满足旁观者的情感需求(Darley, 2002)。也有研究表明，惩罚行为具有能够对违规者形成一种震慑的间接作用，防止这种过激行为的再次发生(Nagin, 1998)。人们通过惩罚可以强调自己与违规者之间的心理距离，以提高他人对自己的社会认同和获得社会赞誉。同时，也有研究者认为第三方对违规者做出的惩罚行为是一种报复性惩罚(Tyler et al., 1997)。在人类社会中，为维护社会准则所做出的第三方惩罚广泛存在着，对此，不同的理论对其发生机制有着不同的理解。

间接互惠理论(Nowak & Sigmund, 1998)依赖于这样一种观点，第三方做出的维护公平准则行为，实际上是为了维护自己的声誉，以便未来与他人互动合作时，这种声誉会给他人传递一种信息，即此人有利他倾向。作为利他惩罚中的一种，第三方惩罚相比于利益直接受的第二方所做出的第二方惩罚，更为中立和客观，因为第三代所处的位置较为特殊，其与违规者和第二方都无关。但从演化的角度出发，该行为与具有适应性的其他一般行为均不一致，对个体而言，用自然选择理论难以对第三方惩罚进行解释。然而，有研究者基于间接互惠理论提出第三方惩罚对于个体而言仍然是具有一定的适应性(Jordan et al., 2016)。即从观察者的角度出发，第三方惩罚的实施者被知觉为富有公德心、值得他人信赖，相应地，违规者则会知觉到第三方惩罚者不是“软柿子”、不可“搭其便车”。因此，从演化角度来说，第三方惩罚对群体适应性和个人声誉均具有积极作用。也就是说，虽然第三方惩罚可能看起来是“利他的”，但从长远来看，由于其他方面所带来的益处，如来自他人更高的声誉和赞扬，这种惩罚很可能是利己主义的。公平偏好理论从社会偏好的角度去解读第三方惩罚，认为人们之所以做出第三方惩罚行为，是因为在事件中有一方做出了不公平的行为，且该行为给其他人在一定程度上造成了损失，是出于对不公平本身的厌恶，才导致人们做出了第三方惩罚行为。该理论认为，现实中存在着相当数量的偏好公平结果的个体，他们对不公平结果的厌恶产生了希望消除这种结果的动机，从而产生了第三方惩罚与传统利己主义模型预测所不同的一系列行为(Fehr & Schmdit, 1999)。社会规范区别于政策法规等硬性规定，是在一个群体中被群体成员所认可并接受和遵从的不成文的行为标准和规则，社会规范激活理论(Schwartz, 1977)则认为社会规范可以通过减少亲个体行为的可能性来简化个体的行为决策并在个体面对复杂、不确定甚至是危险的情境时得到行为上的指引(Pillutla & Chen, 1999; 陈思静等，2015)。国外一批学者最早把第三方惩罚行为与社会规范联系起来，认为第三方惩罚行为是社会规范得以产生、维持和发展的重要原因(Elizabeth et al., 2010)。

3. 第三方惩罚的神经机制研究

在行为研究的基础之上，有研究者对第三方惩罚的生理机制进行了一定的探索。有关第三方惩罚的研究发现，内侧额叶负波(medial frontal negativity, MFN)与第三方的不公平感知有关，高利他主义的第三方，对高不公平分配诱发出更大的 MFN；而低利他主义的第三方，对公平分配将诱发出更大的 MFN (Sun et al., 2015)。第三方对不公平方案做出反应时，将诱发更大的反馈相关负波(feedback-related negative, FRN)；对公平方案做出反应时，将诱发更小的 P300 (Mothes et al., 2016)。第三方对违背社会规范的认知，可能发生在评价过程的早期阶段(Dickson & Wicha, 2019)；而近期有研究者发现，当第三方面对公平方案时将诱发更大的晚期正成分(late positive component, LPC)，面对不公平方案时将诱发较小的 LPC (Cui et al., 2019)。此外，基于第三方惩罚，Chen 等人(2017)的研究表明，FRN 和 P300 与个体的合作行为有关。近期的研究也发现，出现在额叶的 P300 与顶叶的 P300 波幅与合作有关(Zhang et al., 2019)。吴燕和罗跃嘉(2011)通过多次信任博弈任务，探索了对维护合作规范的利他惩罚结果评价的神经机制。结果发现，对利他惩罚的结果进行评价时诱发了 FRN，表明利他惩罚与负性情绪加工有关。

Hu 等人(2015)采用功能磁共振成像(functional magnetic resonance imaging, fMRI)技术，发现第三方惩

罚激活了双侧纹状体外，也激活了左侧前额叶皮层(left prefrontal cortex, LPFC)和腹内侧前额叶皮层(ventromedial prefrontal cortex, vmPFC)。也有研究发现，vmPFC 损伤人群的第三方惩罚要比正常人群少，也进一步说明 vmPFC 是第三方惩罚的关键脑区(Asp et al., 2019)。先前的神经影像学研究表明，第三方惩罚决策是基于两个神经网络：右侧颞顶叶联合区(right temporal parietal junction, rTPJ)的“心理理论网络”和右侧背外侧前额叶皮层(right dorsolateral prefrontal cortex, rDLPFC)的“中央执行网络”(Bellucci et al., 2017)。Strang 等人(2015)研究发现，与第三方惩罚有关的囚徒困境任务中，对右侧 DLPFC 进行抑制将减少了第三方的合作行为。由此可见，维护社会规范的第三方利他惩罚决策活动与 TPJ 和 DLPFC 的活动紧密相关。Krueger 和 Hoffman (2016)根据已有研究，整理出第三方惩罚的神经心理框架：前脑岛(anterior insula, AI)、前扣带回(anterior cingulate cortex, ACC)和杏仁核(amygdala)与个体对违规行为的情绪反应有关，后扣带回(posterior cingulate cortex, PCC)和颞顶联合区(temporo-parietal junction, TPJ)与个体对违规行为的意图判断有关；内侧前额叶皮层(mPFC)能够整合前面两部分信息，判断违规者是否应该受到指责，形成“指责信号”；“指责信号”转变为实际的惩罚行为需要依赖背外侧前额叶皮层(DLPFC)和后顶叶皮层(posterior parietal cortex, PPC)。

然而，通过对前人研究梳理发现，关于第三方惩罚的内在生理机制尚未有明确的、统一的定论，且维护社会规范的第三方惩罚的时程加工特点的研究还相对较少。综上所述，从第三方的角度出发探究维护社会规范的利他惩罚相当有必要。

4. 总结与展望

“不以规矩，不能成方圆”。这句话出自战国•邹•孟轲《孟子•离娄上》，告诉世人做事要有一定的规矩，要遵守社会规范。而破坏社会规范必将受到惩诫，利他惩罚就是这种惩诫的主要方式，是维系社会规范的基石和关键。也许有人质疑，生活中并不会经常出现利他惩罚，因为这种惩罚需要个体牺牲自我利益，并且对群体中的其他人有利。然而，事实上，每个人都有维护社会规范的意愿，绝大多数愿意付出成本去惩罚破坏社会规范的人，比如人们常说的“路见不平、拔刀相助”。

目前，有关第三方惩罚的研究虽然已经逐步深入，但尚存部分重要的问题未被阐明。首先，第三方利他惩罚与日常生活息息相关，而现实生活中人们的行为决策绕不开收益与损失的情境。而当下的研究主要集中在收益情境，已有行为研究表明，在损失情境下第三方愿意实施更多的利他惩罚维护社会公平规范(Liu et al., 2017)，但是得失情境下，产生这种利他惩罚决策差异还亟待探究。此外，现有研究主要聚焦在行为层面考察维护社会规范的第三方利他惩罚的心理机制及其影响因素(Liu et al., 2018)，其内在神经机制仍需进一步研究。最后，维护不同社会规范的第三方惩罚，如公平规范和合作规范，可能是基于不同的心理过程，因此，维护不同规范的第三方惩罚有何异同也值得研究者进行探索。

参考文献

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, (3),* 389-405.

吴燕, 罗跃嘉(2011). 利他惩罚中的结果评价——ERP 研究. *心理学报, 43(6),* 661-673.

Asp, E. W., Gullickson, J. T., Warner, K. A., Koscik, T. R., Denburg, N. L., & Tranel, D. (2019). Soft on Crime: Patients with Ventromedial Prefrontal Cortex Damage Allocate Reduced Third-Party Punishment to Violent Criminals. *Cortex, 119,* 33-45. https://doi.org/10.1016/j.cortex.2019.03.024

Bellucci, G., Chernyak, S., Hoffman, M., Deshpande, G., Dal Monte, O., Knutson, K. M., & Krueger, F. (2017). Effective Connectivity of Brain Regions Underlying Third-Party Punishment: Functional MRI and Granger Causality Evidence. *Social Neuroscience, 12,* 124-134. https://doi.org/10.1080/17470919.2016.1153518

Bendor, J., & Swistak, P. (2001). The Evolution of Norms. *The American Journal of Sociology, 106,* 1493-1545. https://doi.org/10.1086/321298

Chen, Y., Lu, J., Wang, Y., Feng, Z., & Yuan, B. (2017). Social Distance Influences the Outcome Evaluation of Cooperation

and Conflict: Evidence from Event-Related Potentials. *Neuroscience Letters, 647,* 78-84. https://doi.org/10.1016/j.neulet.2017.03.018

Cui, F., Wang, C., Cao, Q., & Jiao, C. (2019). Social Hierarchies in Third-Party Punishment: A Behavioral and ERP Study. *Biological Psychology, 146,* Article ID: 107722. https://doi.org/10.1016/j.biopsycho.2019.107722

Darley, J. (2002). Just Punishments: Research on Retributional Justice. In M. Ross, & D. T. Miller (Eds.), *The Justice Motive in Everyday Life* (pp. 314-333). Cambridge University Press. https://doi.org/10.1017/CBO9780511499975.017

Dickson, D. S., & Wicha, N. Y. (2019). P300 Amplitude and Latency Reflect Arithmetic Skill: An ERP Study of the Problem Size Effect. *Biological Psychology, 148,* Article ID: 107745. https://doi.org/10.1016/j.biopsycho.2019.107745

Elizabeth, T., Antonio, R., & Camerer, F. (2010). Neural Evidence for Inequality-Averse Social Preferences. *Nature, 463,* 1089-1091. https://doi.org/10.1038/nature08785

Fehr, E., & Fischbacher, U. (2004a). Social Norms and Human Cooperation. *Trends in Cognitive Sciences, 8,* 185-190. https://doi.org/10.1016/j.tics.2004.02.007

Fehr, E., & Fischbacher, U. (2004b). Third-Party Punishment and Social Norms. *Evolution and Human Behavior, 25,* 63-87. https://doi.org/10.1016/S1090-5138(04)00005-4

Fehr, E., & Gächter, S. (2002). Altruistic Punishment in Humans. *Nature, 415,* 137-140. https://doi.org/10.1038/415137a

Fehr, E., & Schmdit, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics, 114,* 817-868. https://doi.org/10.1162/003355399556151

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An Experimental Analysis of Ultimatum Bargaining. *Journal of Economic Behavior & Organization, 4,* 367-388. https://doi.org/10.1016/0167-2681(82)90011-7

Hu, Y., Strang, S., & Weber, B. (2015). Helping or Punishing Strangers: Neural Correlates of Altruistic Decisions as Third-Party and of Its Relation to Empathic Concern. *Frontiers in Behavioral Neuroscience, 9,* 24-35. https://doi.org/10.3389/fnbeh.2015.00024

Jordan, J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-Party Punishment as a Costly Signal of Trustworthiness. *Nature, 530,* 473-476. https://doi.org/10.1038/nature16981

Krueger, F., & Hoffman, M. (2016). The Emerging Neuroscience of Third-Party Punishment. *Trends in Neuroscience, 39,* 499-501. https://doi.org/10.1016/j.tins.2016.06.004

Liu, Y. J., Bain, X. H., Hu, Y., Chen, Y. T., Li, X. Z., & Baxter, D. F. (2018). The Influence of Intergroup Bias on Altruistic Behaviors: Ingroup Attenuates Altruistic Punishment. *Social Behavior & Personality: An International Journal, 46,* 1397-1408. https://doi.org/10.2224/sbp.7193

Liu, Y. J., Lin, L., Li, Z., & Guo, X. Y. (2017). Punish the Perpetrator or Compensate the Victim Gain vs. Loss Context Modulate Third-Party Altruistic Behaviors. *Frontiers in Psychology, 8,* 2066-2077. https://doi.org/10.3389/fpsyg.2017.02066

Mothes, H., Enge, S., & Strobel, A. (2016). The Interplay between Feedback-Related Negativity and Individual Differences in Altruistic Punishment: An EEG Study. *Cognitive, Affective, & Behavioral Neuroscience, 16,* 276-288. https://doi.org/10.3758/s13415-015-0388-x

Nagin, D. (1998). Deterrence and Incapacitation. In M. Tonry (Ed.), *The Handbook of Crime and Punishment* (pp. 345-368). Oxford University Press.

Nowak, M. A., & Sigmund, K. (1998). Evolution of Indirect Reciprocity by Image Scoring. *Nature, 393,* 573-577. https://doi.org/10.1038/31225

Pillutla, M. M., & Chen, X. P. (1999). Social Norms and Cooperation in Social Dilemmas: The Effects of Context and Feedback. *Organizational Behavior and Human Decision Processes, 78,* 81-103. https://doi.org/10.1006/obhd.1999.2825

Schwartz, S. H. (1977). Normative Influence on Altruism. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10, pp. 221-279). Academic Press. https://doi.org/10.1016/S0065-2601(08)60358-5

Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B., & Sack, A. T. (2015). Be Nice If You Have to—The Neurobiological Roots of Strategic Fairness. *Social Cognitive and Affective Neuroscience, 10,* 790-796. https://doi.org/10.1093/scan/nsu114

Sun, L., Tan, P., Cheng, Y., Chen, J., & Qu, C. (2015). The Effect of Altruistic Tendency on Fairness in Third-Party Punishment. *Frontiers in Psychology, 6,* 820-830. https://doi.org/10.3389/fpsyg.2015.00820

Tyler, T. R., Boeckmann, R. J., Smith, H. J., & Huo, Y. J. (1997). Social Justice in a Diverse Society. *American Political Science Association, 92,* 839-842.

Zhang, D., Lin, Y., Jing, Y., Feng, C., & Gu, R. (2019). The Dynamics of Belief Updating in Human Cooperation: Findings from Inter-Brain ERP Hyperscanning. *NeuroImage, 198,* 1-12. https://doi.org/10.1016/j.neuroimage.2019.05.029