Advances in Psychology 心理学进展, 2021, 11(7), 1644-1652
Published Online July 2021 in Hans. http://www.hanspub.org/journal/ap
https://doi.org/10.12677/ap.2021.117184

Hans汉斯

第二方与第三方惩罚的发生机制及其差异

马德鑫，曹晓君

西华师范大学教育学院，四川 南充
Email: cxj841118@aliyun.com

收稿日期：2021年6月1日；录用日期：2021年6月25日；发布日期：2021年7月6日

摘　要

利他惩罚行为是指需要个体牺牲自己利益的，有利于群体中大多数人的，对违反社会规范的人做出的惩罚行为。利他惩罚分为第二方惩罚与第三方惩罚。通过分析两者的神经机制、作用途径和效果，并结合相关理论及模型。发现两者均由情绪系统与社会认知系统共同控制，但第三方惩罚中有更多的认知系统参与。第三方惩罚也比第二方惩罚更加有效。

关键词

利他惩罚，第二方惩罚，第三方惩罚，效果

The Mechanism of Second-Party and Third-Party Punishment and Its Differences

Dexin Ma, Xiaojun Cao

Department of Education, China West Normal University, Nanchong Sichuan
Email: cxj841118@aliyun.com

Received: Jun. 1st, 2021; accepted: Jun. 25th, 2021; published: Jul. 6th, 2021

Abstract

Altruistic punishment is a punitive act that requires individuals to sacrifice their own interests in favor of the majority of people in the group, and is made against those who violate social norms. Altruistic punishment is divided into second-party punishment and third-party punishment. By analyzing the neural mechanisms, action pathways and effects of both, and combining with related theories and models. Both are found to be controlled by the emotional system and the social cognitive system, but more cognitive systems are involved in third-party punishment. Third-party punishment is also more effective than second-party punishment.

文章引用: 马德鑫, 曹晓君(2021). 第二方与第三方惩罚的发生机制及其差异. *心理学进展, 11(7),* 1644-1652.
DOI: 10.12677/ap.2021.117184

Keywords

Altruistic Punishment, Second-Party Punishment, Third-Party Punishment, Effects

1. 引言

在日常生活中，有的人会见义勇为，“路见不平，拔刀相助”。像这样的，需要个体牺牲自己利益，有利于群体中大多数人的，对违反社会规范的人做出的惩罚行为被称为利他惩罚(Altruistic Punishment) (Boyd, Gintis, Bowles, & Richerson, 2003; Fehr & Fischbacher, 2003)。利他惩罚又分为第二方惩罚(Second Party Punishment, SPP)与第三方惩罚(Third Party Punishment, TPP)。当一方违反社会规范时，利益受损的另一方(自我利益卷入方)会对违规者采取惩罚措施来维护其他人的利益，这种行为被称为第二方惩罚(Fehr & Ga, 2002)；当一方违反社会规范的行为，自身利益无损的另一方牺牲自我利益来惩罚违规者，从而维护其他人利益，这种行为被称为第三方惩罚(Fehr & Fischbacher, 2004)。两者的区别在于第二方利他惩罚者处于利益受损状态，第三方利他惩罚者处于旁观者地位(Gummerum, López-Pérez, Van Dijk, & Van Dillen, 2020; McAuliffe, Jordan, &Warneken, 2015)。

人类社会依赖于其成员的合作行为(Seier, 2020)，利他惩罚作为违规行为的反制手段，能够有效促进合作行为，阻止搭便车行为，维护社会规范(Bond, 2019; Boyd, Gintis, Bowles, & Richerson, 2003; Fehr & Ga, 2002)。通过研究利他惩罚可以从侧面解释人类社会中常见的合作行为。但利他惩罚不是一种理性的行为(Gintis, Bowles, Boyd, & Fehr, 2003; Uchida, Yamamoto, Okada, & Sasaki, 2019)，惩罚需要付出成本，“理性”的人不会选择去惩罚违规者。一些研究者试图从不同角度解释利他惩罚，并提出了相关理论。强互惠理论(Strong Reciprocity Theory)从群体利益角度(Bowles & Gintis, 2004)出发，高成本信号模型(Costly Singaling Model)则根据声誉和个人可信度来解释(Jordan, Hoffman, Bloom, & Rand, 2016)，从遵守规范的角度也能解释利他惩罚(Kimbrough & Vostroknutov, 2015)。

一方面，利他惩罚作为亲社会行为，研究其影响因素与神经机制是有必要的。另一方面，利他惩罚又作为一种对违规行为的反制手段，还需要关注利他惩罚的效果。本文主要梳理相关文献，系统讨论第二方与第三方惩罚的产生机制与效果，为理解利他惩罚提供一个新的视角。

2. 第三方惩罚更“理性”，第二方惩罚更“感性”

情绪在社会决策中扮演着一个很重要的角色，也是影响利他惩罚的一大影响因素(Engelmann & Fehr, 2017; Koppel, Andersson, Västfjäll, & Tinghög, 2017)。消极情绪会增加负反馈行为(Drouvelis & Grosskopf, 2016; Lane, 2017)，促进第二方惩罚(Fehr & Ga, 2002)。惩罚行为被认为是受到情绪和认知的影响(Egas & Riedl, 2008)。情绪系统与认知系统共同控制了第二方与第三方惩罚，但在第三方惩罚中以认知系统为主导，在第二方惩罚中以情绪系统为主导。

在利他惩罚的神经机制方面，分级惩罚模型认为惩罚分为三类(Krueger & Hoffman, 2016)。第一类为“第一方惩罚”，即良知和内疚，第二方惩罚起源于第一方惩罚，而第三方惩罚又是第二方惩罚的扩展。第二方与第三方惩罚激活了大量相同的脑区(Strobel et al., 2011)，包括背外侧前额皮层(Dorsolateral Pre-

frontal Cortex)和双侧前脑岛(Bilateral Anterior Insula)。但在功能上，社会认知区域(颞顶交界处)参与了第三方惩罚(Bellucci, Camilleri, Iyengar, Eickhoff, & Krueger, 2020)。认知控制相关脑区受损的参与者在第三方惩罚方面异于常人(Glass, Moody, Grafman, & Krueger, 2016)。例如，腹内侧脑前额叶(Ventromedial Prefrontal Cortex)损伤的患者会减轻对第三方罪犯的惩罚(Asp et al., 2019)，右侧背外侧前额叶区域的活动也会调节第三方利他惩罚行为(殷西乐，李建标，陈思宇，刘晓丽，郝洁，2019)。但第二方惩罚更多由情绪相关功能的脑区控制(Bellucci, Camilleri, Iyengar, Eickhoff, & Krueger, 2020)，杏仁核也对第二方惩罚有调节作用(Scheele et al., 2012)。

也有一些实验证据可从侧面论证第三方惩罚更“理性”。事实证明，利他惩罚并不是一种理性的行为(Mussel, Hewig, & Weiss, 2018; Uchida, Yamamoto, Okada, & Sasaki, 2019)。一个人自我控制状态对第二方惩罚有影响，自我损耗促进了第二方惩罚(Crockett, Clark, Lieberman, Tabibnia, & Robbins, 2010; Liu, He, & Dou, 2015)。但是在第三方惩罚中，认知控制可以帮助个体抑制自利倾向，自我控制水平可以预测规范执行行为(Friehe & Schildberg-Hörisch, 2018)，避免因为成本不惩罚违规者(谢东杰，苏彦捷，2019)。

所以，相比较于第二方惩罚，第三方惩罚更多地由社会认知系统控制，显得更加理性。

3. 第二方与第三方惩罚的效果

在生活中，利他惩罚可以促进合作，维护规范。在实验中，利他惩罚的作用主要通过群体内合作行为的增加或搭便车行为的减少来体现(Bond, 2019; Fehr & Ga, 2002; Waichman & Stenzel, 2019)。但利他惩罚对不同对象(违规者、惩罚者、群体等)的作用机制或具体效果是不同的。

利他惩罚对违规者的作用主要是通过威慑机制(Krasnow, Delton, Cosmides, & Tooby, 2016)。威慑理论认为惩罚者通过惩罚来威慑违规者，防止违规者再次违规。违规者在受到惩罚后，会因为害怕再次受到惩罚而增加合作行为。

利他惩罚对群体的作用一方面是通过声誉机制产生。高成本信号模型认为人可以分为两类(Jordan, Hoffman, Bloom, & Rand, 2016)。一类是值得信赖的人，一类是自私的人。高成本信号模型假设，“如果你看见我惩罚违规者，我会被你标记为一个值得信赖的人”。在利他惩罚后，他人会认为惩罚者是一个值得信赖的人，从而惩罚者在群体中获得一定声誉。而声誉在随后的互动中会带来一定的利益(Barclay, 2006; Jordan & Rand, 2017a; Raihani & Bshary, 2015)。另外，利他惩罚也可以威慑群体中的其他人，防止潜在的违规者出现(Delton & Krasnow, 2017; Krasnow, Delton, Cosmides, & Tooby, 2016)。最后，利他惩罚可以在群体内中形成和维护一定的规范。例如，公平是人类社会的基本社会规范(郭秀艳，郑丽，程雪梅，刘映杰，李林，2017)，利他惩罚的对象往往就是不公平的分配者(Fehr & Ga, 2002; Gummerum, López-Pérez, Van Dijk, & Van Dillen, 2020; McAuliffe & Dunham, 2017; Seier, 2020)，由此维护了公平规范。通过这些途径，利他惩罚维护了群体的利益(Chaudhuri & Paichayontvijit, 2017; Gächter, Renner, & Sefton, 2008)，促进了群体的发展。

利他惩罚对惩罚者本身也有一定的作用。首先，利他惩罚可以平复惩罚者的负面情绪。个体在面对不道德的行为时，会产生愤怒、厌恶等负面情绪(Jordan, Mcauliffe, & Rand, 2016; Salerno & Peter-Hagene, 2013)，甚至接触不道德物品也会使人产生负面情绪(Eskine, Novreske, & Richards, 2013)。而利他惩罚激活了大脑的奖赏回路(de Quervain et al., 2004; Strobel et al., 2011)，有效的利他惩罚会使个体获得满足感与正面激励(Mussel, Hewig, & Weiss, 2018)。正如利他行为有助于个体获得正性情绪一样(Hu, Li, Jia, & Xie, 2016)，惩罚者从利他惩罚中获得了快乐(Yamagishi et al., 2017)。其次，利他惩罚也可以在未来为惩罚者获得更多的利益。利他惩罚可以帮助个体获得一定的声誉(Jordan, Hoffman, Bloom, & Rand, 2016; Raihani

& Bshary, 2015)和威慑力(Krasnow, Delton, Cosmides, & Tooby, 2016)，惩罚者在之后的互动中更容易被选为同伴(Jordan & Rand, 2017b)，并且惩罚的成本越多越容易被选为合作伙伴(Nelissen, 2008)。

在此，通过分析利他惩罚的神经机制与效果，并结合信息加工的观点把第二方与第三方惩罚整合到一个理论框架中，如图 1 所示。

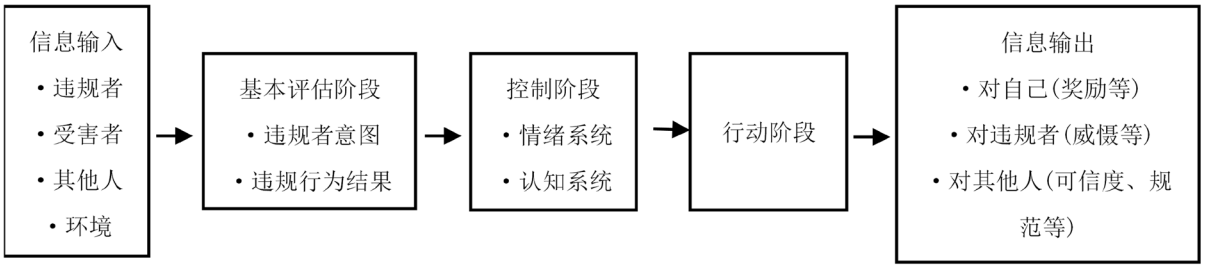

Figure 1. The theoretic model of the second-party & third-party punishment
图 1. 第二方与第三方惩罚的理论模型

两者基本的流程是一样的，都经历信息输入、基本评估阶段、控制阶段、反应阶段、信息输出等过程。但在控制阶段，第三方惩罚更多地由认知系统控制，第二方惩罚更多地由情绪系统控制。此外，经过在不同阶段的对比，总的发现第三方惩罚比第二方惩罚更有效。以下为具体的分析与阐述。

4. 第三方惩罚比第二方惩罚更有效

从性质上看，第二方惩罚与第三方惩罚都被认为是一种特殊的亲社会行为(Gintis, Bowles, Boyd, & Fehr, 2003)。两者的目的都是为了维护社会规范，第二方与第三方惩罚者都倾向对违反社会规范的人实施制裁。但因两者的神经机制与造成的具体影响不同，总的来说第三方惩罚比第二方惩罚更有效。

第三方惩罚需要的信息更多，向外传递的信息也更多。在第二方情境中，参与交互的人数最少可以为 2 人。而在第三方情境中，参与交互的至少为 3 人，这样惩罚者需要考虑的信息会更多。第二方惩罚者自身就受到了侵犯，其关注点主要在违规判断、情绪反应及伤害大小上。第三方惩罚者还关注到责任的评估(Bellucci, Camilleri, Iyengar, Eickhoff, & Krueger, 2020; Krueger & Hoffman, 2016)及成本。第二方惩罚者可以发出一种信号，即自己受到侵犯时会反击，以此来威慑那些潜在的侵犯者。第三方惩罚者也会发出这种信号威慑他人(Krasnow, Delton, Cosmides, & Tooby, 2016)。惩罚者通过实施第三方惩罚，获得一定的声誉，增加自己的可信度。而声誉在随后的活动中会带来一定的好处(Barclay, 2006; Jordan & Rand, 2017a; Raihani & Bshary, 2015)。第二方惩罚并不能充分展示自己是个值得信赖的人，在分享行为存在的条件下，第二方惩罚甚至会降低信任感(Przepiorka & Liebe, 2016)。第二方惩罚相比，第三方惩罚传达了更多积极信息，更能引导人们减少自私行为(Chen, Zeng, & Ma, 2020)。

另外，惩罚者的动机往往受到怀疑。惩罚者可能由于报复、不平等厌恶、预防违规等动机而实施利他惩罚(Bone & Raihani, 2015; Raihani & McAuliffe, 2012; Twardawski & Hilbig, 2020)。但有的研究者认为第二方惩罚主要的动机是报复(Deutchman, Bracic, Raihani, & McAuliffe, 2020)，是一种反抗的手段(Yamagishi et al., 2012)。

第三方惩罚也更适合人类社会。相对于其他物种，只有人类具有第三方惩罚的能力(Leimgruber, Rosati, & Santos, 2016; Riedl, Jensen, Call, & Tomasello, 2012)，第三方惩罚需要儿童心理理论能力成熟后才出现(Bellucci, Camilleri, Iyengar, Eickhoff, & Krueger, 2020)。强互惠理论认为实施利他惩罚不能直接增加个体的适应性。实施利他惩罚是通过惩罚这种手段来维护合作或者社会规范来增加群体的适应性(Bowles &

Gintis, 2004)。但第二方惩罚影响较小，第三方惩罚是维护大规模社会规范的关键(Fehr & Fischbacher, 2004)。也有研究证实，在更大、更复杂的社会中人们遭受的第三方惩罚要多得多(Marlowe et al., 2007)。

最后，违规者可能由于认知偏见而使第三方惩罚更加有效。第三方惩罚在实验室中有跨文化一致性(Bernhard, Fischbacher, & Fehr, 2006)和普遍性，但是在日常生活中第三方惩罚可能要稀少得多(Balafoutas, Nikiforakis, & Rockenbach, 2014, 2016; Pedersen et al., 2020)。前景理论认为人们会基于参照点而改变对风险的态度。在现实生活中，被惩罚者往往将小概率和损失估计的比实际更大，觉得惩罚比实际情况更严厉、更频繁(Uchida, Yamamoto, Okada, & Sasaki, 2019)。

5. 研究展望

5.1. 利他惩罚的神经机制及发展趋势需要进一步探究

虽然在动物界种内和种间的利他行为有很多例子(Herzing, 2014)，但是只有在一些灵长类动物中发现有利他惩罚行为(Leimgruber, Rosati, & Santos, 2016)，且发现黑猩猩没有第三方惩罚(Riedl, Jensen, Call, & Tomasello, 2012)，这表明似乎只有人类具有第三方惩罚的能力。然而人类中广泛的存在第三方惩罚行为。有什么因素能够促进灵长类动物发展出第二方惩罚行为和人类发展出第三方惩罚行为也需要进行更加深入的研究。

个体生命全程中利他惩罚的发展趋势及利他惩罚形成的关键期等一系列问题也尚未解决。例如，为什么青少年时期表现出更多的利他惩罚？与青少年时期亲社会行为快速发展还是情绪情感较敏感有关呢？在最后通牒博弈中，老年人更容易拒绝中等程度的不公平提议，左外侧前额皮层表现出更大的激活，双侧前脑岛也表现出更小的激活(Harlé & Sanfey, 2012)。这表明进入老年期有可能会对个体的利他惩罚行为产生影响。目前老年人在社会中所占的比例越来越多(李建伟，周灵灵，2018)，理解老年人的利他惩罚发展情况及影响因素，可以更有效的运用利他惩罚这一机制促进老年群体的合作行为及维护社会规范。

5.2. 从报复的角度来加深对两种利他惩罚的理解

报复指在受到伤害后，有意识地对侵犯者进行伤害或者施加痛苦(Schumann & Ross, 2010)。在道德和理性程度上，报复行为是最低的，其次是第二方惩罚。Scheele 等人认为第二方惩罚是具有报复冲动的一种行为(Scheele et al., 2012)。或许利他惩罚是报复行为的一种升华，有研究表示报复是一种情感上的优势反应(Will, Crone, Van Lier, & Güroğlu, 2016)，而利他惩罚需要一定的自我控制能力和公平道德判断(Friehe & Schildberg-Hörisch, 2018)。另外，报复行为与利他惩罚密切相关，尤其是第二方惩罚，在影响因素和神经机制上有很多相似的地方(Fessler, 2006; Jackson, Choi, & Gelfand, 2019; Krueger & Hoffman, 2016; Sindermann et al., 2018; Zourrig, Chebat, & Toffoli, 2015)。在神经机制上，报复与利他惩罚都激活了与奖励活动相关的脑区(Chester & DeWall, 2015; Strobel et al., 2011)。愤怒情绪与报复密切相关，可以有效预测报复行为的产生(Jackson, Choi, & Gelfand, 2019)，同样也可以增加利他惩罚。报复在进化角度上先于第二、三方惩罚出现(Akçay et al., 2009; Silk, 2005)，非灵长类动物也有报复行为。报复、第二方利他惩罚、第三方利他惩罚似乎在进化上是一脉相承的。或许可以在以后的利他惩罚发展研究中，从报复的角度理解利他惩罚。

5.3. 通过整合多方因素探讨两种利他惩罚的作用机制

利他惩罚不单单是一种心理现象和行为，利他惩罚也可以作为一种促进合作的工具。认识第二方和第三方惩罚不仅要评析这两者的影响因素，也要探析如何使两者更好地促进合作及维护规范。比如相对收益(即实施惩罚的成本与效果)在利他惩罚中也起到非常大的作用，会影响利他惩罚对促进合作的效果

(Egas & Riedl, 2008; Nikiforakis & Normann, 2008)。未来也可以研究如何扩大利他惩罚的效用及如何更好地利用这一工具。

个体差异及他人对利他惩罚的影响也有待进一步探究。例如共情对利他惩罚的影响有相互矛盾的结果(Lu & McKeown, 2018; Weng, Fox, Hessenthaler, Stodola, & Davidson, 2015)。在不同的文化中，人们如何看待惩罚违背社会规范的这种行为也有差异(Eriksson et al., 2017)。因此在不同的文化中，如何促进个体的利他惩罚，如何更好利用利他惩罚促进合作和维护规范也需要更多的研究。

参考文献

郭秀艳, 郑丽, 程雪梅, 刘映杰, 李林(2017). 不公平感及相关决策的认知神经机制. *心理科学进展, 25(6),* 903-911.

李建伟, 周灵灵(2018). 中国人口政策与人口结构及其未来发展趋势. *经济学动态, (12),* 4.

谢东杰, 苏彦捷(2019). 第三方惩罚的演化与认知机制. *心理科学, 42(1),* 216-222.

殷西乐, 李建标, 陈思宇, 刘晓丽, 郝洁(2019). 第三方惩罚的神经机制: 来自经颅直流电刺激的证据. *心理学报, 51(5),* 49-61.

Akçay, Ç., Wood, W. E., Searcy, W. A., Templeton, C. N., Campbell, S. E., & Beecher, M. D. (2009). Good Neighbour, Bad Neighbour: Song Sparrows Retaliate against Aggressive Rivals. *Animal Behaviour*, *78,* 97-102. https://doi.org/10.1016/j.anbehav.2009.03.023

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

Balafoutas, L., Nikiforakis, N., & Rockenbach, B. (2014). Direct and Indirect Punishment among Strangers in the Field. *Proceedings of the National Academy of Sciences, 111,* 15924-15927. https://doi.org/10.1073/pnas.1413170111

Balafoutas, L., Nikiforakis, N., & Rockenbach, B. (2016). Altruistic Punishment Does Not Increase with the Severity of Norm Violations in the Field. *Nature Communications, 7,* Article No. 13327. https://doi.org/10.1038/ncomms13327

Barclay, P. (2006). Reputational Benefits for Altruistic Punishment. *Evolution and Human Behavior, 27,* 325-344. https://doi.org/10.1016/j.evolhumbehav.2006.01.003

Bellucci, G., Camilleri, J. A., Iyengar, V., Eickhoff, S. B., & Krueger, F. (2020). The Emerging Neuroscience of Social Punishment: Meta-Analytic Evidence. *Neuroscience & Biobehavioral Reviews, 113,* 426-439. https://doi.org/10.1016/j.neubiorev.2020.04.011

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial Altruism in Humans. *Nature, 442,* 912-915. https://doi.org/10.1038/nature04981

Bond, R. M. (2019). Low-Cost, High-Impact Altruistic Punishment Promotes Cooperation Cascades in Human Social Networks. *Scientific Reports, 9,* Article No. 2061. https://doi.org/10.1038/s41598-018-38323-7

Bone, J., & Raihani, N. (2015). Human Punishment Is Motivated by Both a Desire for Revenge and a Desire for Equality. *Evolution and Human Behavior, 36,* 323-330. https://doi.org/10.1016/j.evolhumbehav.2015.02.002

Bowles, S., & Gintis, H. (2004). The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations. *Theoretical Population Biology, 65,* 17-28. https://doi.org/10.1016/j.tpb.2003.07.001

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The Evolution of Altruistic Punishment. *Proceedings of the National Academy of Sciences, 100,* 3531-3535. https://doi.org/10.1073/pnas.0630443100

Chaudhuri, A., & Paichayontvijit, T. (2017). On the Long-Run Efficacy of Punishments and Recommendations in a Laboratory Public Goods Game. *Scientific Reports, 7,* Article No. 12286. https://doi.org/10.1038/s41598-017-12490-5

Chen, H., Zeng, Z., & Ma, J. (2020). The Source of Punishment Matters: Third-Party Punishment Restrains Observers from Selfish Behaviors Better than Does Second-Party Punishment by Shaping Norm Perceptions. *PLoS ONE, 15,* e0229510. https://doi.org/10.1371/journal.pone.0229510

Chester, D. S., & DeWall, C. N. (2015). The Pleasure of Revenge: Retaliatory Aggression Arises from a Neural Imbalance toward Reward. *Social Cognitive and Affective Neuroscience, 11,* 1173-1182. https://doi.org/10.1093/scan/nsv082

Crockett, M. J., Clark, L., Lieberman, M. D., Tabibnia, G., & Robbins, T. W. (2010). Impulsive Choice and Altruistic Punishment Are Correlated and Increase in Tandem with Serotonin Depletion. *Emotion, 10,* 855. https://doi.org/10.1037/a0019861

de Quervain, D. D., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The Neural Basis of Altruistic Punishment. *Science, 305,* 1254-1258. https://doi.org/10.1126/science.1100735

Delton, A. W., & Krasnow, M. M. (2017). The Psychology of Deterrence Explains why Group Membership Matters for Third-Party Punishment. *Evolution and Human Behavior, 38,* 734-743. https://doi.org/10.1016/j.evolhumbehav.2017.07.003

Deutchman, P., Bracic, M., Raihani, N., & McAuliffe, K. (2020). Punishment Is Strongly Motivated by Revenge and Weakly Motivated by Inequity Aversion. *Evolution and Human Behavior, 42,* 12-20. https://doi.org/10.1016/j.evolhumbehav.2020.06.001

Drouvelis, M., & Grosskopf, B. (2016). The Effects of Induced Emotions on Pro-Social Behaviour. *Journal of Public Economics, 134,* 1-8. https://doi.org/10.1016/j.jpubeco.2015.12.012

Egas, M., & Riedl, A. (2008). The Economics of Altruistic Punishment and the Maintenance of Cooperation. *Proceedings of the Royal Society B: Biological Sciences, 275,* 871-878. https://doi.org/10.1098/rspb.2007.1558

Engelmann, J. B., & Fehr, E. (2017). *The Neurobiology of Trust and Social Decision-Making: The Important Role of Emotions*. https://doi.org/10.1093/oso/9780190630782.003.0003

Eriksson, K., Strimling, P., Andersson, P. A., Aveyard, M., Brauer, M., Gritskov, V., Kiyonari, T., Kuhlman, D. M., Maitner, A. T., Manesi, Z. et al. (2017). Cultural Universals and Cultural Differences in Meta-Norms about Peer Punishment. *Management and Organization Review, 13,* 851-870. https://doi.org/10.1017/mor.2017.42

Eskine, K., Novreske, A., & Richards, M. (2013). Moral Contagion Effects in Everyday Interpersonal Encounters. *Journal of Experimental Social Psychology, 49,* 947-950. https://doi.org/10.1016/j.jesp.2013.04.009

Fehr, E., & Fischbacher, U. (2003). The Nature of Human Altruism. *Nature, 425,* 785. https://doi.org/10.1038/nature02043

Fehr, E., & Fischbacher, U. (2004). 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., & Ga, S. (2002). Altruistic Punishment in Humans. *Nature, 415,* 137-140. https://doi.org/10.1038/415137a

Fessler, D. M. T. (2006). The Male Flash of Anger: Violent Response to Transgression as an Example of the Intersection of Evolved Psychology and Culture. In J. H. Barkow (Ed.), *Missing the Revolution: Darwinism for Social Scientists* (pp. 101-117). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195130027.003.0003

Friehe, T., & Schildberg-Hörisch, H. (2018). Predicting Norm Enforcement: The Individual and Joint Predictive Power of Economic Preferences, Personality, and Self-Control. *European Journal of Law and Economics, 45,* 127-146. https://doi.org/10.1007/s10657-017-9556-5

Gächter, S., Renner, E., & Sefton, M. (2008). The Long-Run Benefits of Punishment. *Science, 322,* 1510. https://doi.org/10.1126/science.1164744

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining Altruistic Behavior in Humans. *Evolution and Human Behavior, 24,* 153-172. https://doi.org/10.1016/S1090-5138(02)00157-5

Glass, L., Moody, L., Grafman, J., & Krueger, F. (2016). Neural Signatures of Third-Party Punishment: Evidence from Penetrating Traumatic Brain Injury. *Social Cognitive and Affective Neuroscience, 11,* 253-262. https://doi.org/10.1093/scan/nsv105

Gummerum, M., López-Pérez, B., Van Dijk, E., & Van Dillen, L. F. (2020). When Punishment is Emotion-Driven: Children's, Adolescents', and Adults' Costly Punishment of Unfair Allocations. *Social Development, 29,* 126-142. https://doi.org/10.1111/sode.12387

Harlé, K. M., & Sanfey, A. G. (2012). Social Economic Decision-Making across the Lifespan: An fMRI Investigation. *Neuropsychologia, 50,* 1416-1424. https://doi.org/10.1016/j.neuropsychologia.2012.02.026

Herzing, D. L. (2014). Interspecies Altruism: Learning from Species on Earth. In *Extraterrestrial Altruism* (pp. 191-207). Berlin: Springer. https://doi.org/10.1007/978-3-642-37750-1_12

Hu, T.-Y., Li, J., Jia, H., & Xie, X. (2016). Helping Others, Warming Yourself: Altruistic Behaviors Increase Warmth Feelings of the Ambient Environment. *Frontiers in Psychology, 7,* Article No. 1349. https://doi.org/10.3389/fpsyg.2016.01349

Jackson, J. C., Choi, V. K., & Gelfand, M. J. (2019). Revenge: A Multilevel Review and Synthesis. *Annual Review of Psychology, 70,* 319-345. https://doi.org/10.1146/annurev-psych-010418-103305

Jordan, J. J., & Rand, D. G. (2017a). Foundations of Costly Signaling: Individuals Who Benefit More from Trustworthiness in Daily Life Find Third-Party Punishment Less Costly. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3020984

Jordan, J. J., & Rand, D. G. (2017b). Third-Party Punishment as a Costly Signal of High Continuation Probabilities in Repeated Games. *Journal of Theoretical Biology, 421,* 189-202. https://doi.org/10.1016/j.jtbi.2017.04.004

Jordan, J. 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

Jordan, J., Mcauliffe, K., & Rand, D. (2016). The Effects of Endowment Size and Strategy Method on Third Party Punish-

ment. *Experimental Economics, 19,* 741-763. https://doi.org/10.1007/s10683-015-9466-8

Kimbrough, E. O., & Vostroknutov, A. (2015). Norms Make Preferences Social. *Journal of the European Economic Association, 14,* 608-638. https://doi.org/10.1111/jeea.12152

Koppel, L., Andersson, D., Västfjäll, D., & Tinghög, G. (2017). The (Null) Effect of Affective Touch on Betrayal Aversion, Altruism, and Risk Taking. *Frontiers in Behavioral Neuroscience, 11,* 251. https://doi.org/10.3389/fnbeh.2017.00251

Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2016). Looking under the Hood of Third-Party Punishment Reveals Design for Personal Benefit. *Psychological Science, 27,* 405-418. https://doi.org/10.1177/0956797615624469

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

Lane, T. (2017). How Does Happiness Relate to Economic Behaviour? A Review of the Literature. *Journal of Behavioral and Experimental Economics, 68,* 62-78. https://doi.org/10.1016/j.socec.2017.04.001

Leimgruber, K. L., Rosati, A. G., & Santos, L. R. (2016). Capuchin Monkeys Punish Those Who Have More. *Evolution and Human Behavior*, *37,* 236-244. https://doi.org/10.1016/j.evolhumbehav.2015.12.002

Liu, Y., He, N., & Dou, K. (2015). Ego-Depletion Promotes Altruistic Punishment. *Open Journal of Social Sciences, 3,* 62-69. https://doi.org/10.4236/jss.2015.311009

Lu, T., & McKeown, S. (2018). The Effects of Empathy, Perceived Injustice and Group Identity on Altruistic Preferences: Towards Compensation or Punishment. *Journal of Applied Social Psychology, 48,* 683-691. https://doi.org/10.1111/jasp.12558

Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Ensminger, J., Gurven, M., Gwako, E., Henrich, J. et al. (2007). More "Altruistic" Punishment in Larger Societies. *Proceedings of the Royal Society B: Biological Sciences, 275,* 587-592. https://doi.org/10.1098/rspb.2007.1517

McAuliffe, K., & Dunham, Y. (2017). Fairness Overrides Group Bias in Children's Second-Party Punishment. *Journal of Experimental Psychology: General, 146,* 485. https://doi.org/10.1037/xge0000244

McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly Third-Party Punishment in Young Children. *Cognition, 134,* 1-10. https://doi.org/10.1016/j.cognition.2014.08.013

Mussel, P., Hewig, J., & Weiss, M. A. (2018). The Reward-Like Nature of Social Cues That Indicate Successful Altruistic Punishment. *Psychophysiology, 55,* e13093. https://doi.org/10.1111/psyp.13093

Nelissen, R. M. A. (2008). The Price You Pay: Cost-Dependent Reputation Effects of Altruistic Punishment. *Evolution and Human Behavior, 29,* 242-248. https://doi.org/10.1016/j.evolhumbehav.2008.01.001

Nikiforakis, N., & Normann, H.-T. (2008). A Comparative Statics Analysis of Punishment in Public-Good Experiments. *Experimental Economics, 11,* 358-369. https://doi.org/10.1007/s10683-007-9171-3

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,* 846-853. https://doi.org/10.1177/1948550619884565

Przepiorka, W., & Liebe, U. (2016). Generosity Is a Sign of Trustworthiness—The Punishment of Selfishness Is Not. *Evolution and Human Behavior, 37,* 255-262. https://doi.org/10.1016/j.evolhumbehav.2015.12.003

Raihani, N. J., & Bshary, R. (2015). The Reputation of Punishers. *Trends in Ecology and Evolution, 30,* 98-103. https://doi.org/10.1016/j.tree.2014.12.003

Raihani, N. J., & McAuliffe, K. (2012). Human Punishment Is Motivated by Inequity Aversion, Not a Desire for Reciprocity. *Biology Letters, 8,* 802-804. https://doi.org/10.1098/rsbl.2012.0470

Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2012). No Third-Party Punishment in Chimpanzees. *Proceedings of the National Academy of Sciences, 109,* 14824-14829. https://doi.org/10.1073/pnas.1203179109

Salerno, J., & Peter-Hagene, L. C. (2013). The Interactive Effect of Anger and Disgust on Moral Outrage and Judgments. *Psychological Science, 24,* 2069-2078. https://doi.org/10.1177/0956797613486988

Scheele, D., Mihov, Y., Kendrick, K. M., Feinstein, J. S., Reich, H., Maier, W., & Hurlemann, R. (2012). Amygdala Lesion Profoundly Alters Altruistic Punishment. *Biological Psychiatry, 72,* e5-e7. https://doi.org/10.1016/j.biopsych.2012.01.028

Schumann, K., & Ross, M. (2010). The Benefits, Costs, and Paradox of Revenge. *Social and Personality Psychology Compass, 4,* 1193-1205. https://doi.org/10.1111/j.1751-9004.2010.00322.x

Seier, M. (2020). The Intuition of Punishment: A Study of Fairness Preferences and Cognitive Ability. *Games, 11,* 21. https://doi.org/10.3390/g11020021

Silk, J. B. (2005). The Evolution of Cooperation in Primate Groups. In H. Gintis, S. Bowles, R. Boyd, & E. Fehr (Eds.), *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life* (p. 17). Cambridge, MA: MIT Press.

Sindermann, C., Luo, R., Zhao, Z., Li, Q., Li, M., Kendrick, K. M., Panksepp, J., & Montag, C. (2018). High ANGER and Low Agreeableness Predict Vengefulness in German and Chinese Participants. *Personality and Individual Differences, 121,* 184-192. https://doi.org/10.1016/j.paid.2017.09.004

Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S., & Kirsch, P. (2011). Beyond Revenge: Neural and Genetic Bases of Altruistic Punishment. *Neuroimage, 54,* 671-680. https://doi.org/10.1016/j.neuroimage.2010.07.051

Twardawski, M., & Hilbig, B. E. (2020). The Motivational Basis of Third-Party Punishment in Children. *PLoS ONE, 15,* e0241919. https://doi.org/10.1371/journal.pone.0241919

Uchida, S., Yamamoto, H., Okada, I., & Sasaki, T. (2019). Evolution of Cooperation with Peer Punishment under Prospect Theory. *Games, 10,* 1-14. https://doi.org/10.3390/g10010011

Waichman, I., & Stenzel, L. (2019). When Punishment Strikes Late: The Effect of a Delay in Punishment and Punishment Feedback on Cooperation and Efficiency. *Journal of Neuroscience, Psychology, and Economics, 12,* 1. https://doi.org/10.1037/npe0000099

Weng, H. Y., Fox, A. S., Hessenthaler, H. C., Stodola, D. E., & Davidson, R. J. (2015). The Role of Compassion in Altruistic Helping and Punishment Behavior. *PLoS ONE, 10,* e0143794. https://doi.org/10.1371/journal.pone.0143794

Will, G.-J., Crone, E. A., Van Lier, P. A. C., & Güroğlu, B. (2016). Neural Correlates of Retaliatory and Prosocial Reactions to Social Exclusion: Associations with Chronic Peer Rejection. *Developmental Cognitive Neuroscience, 19,* 288-297. https://doi.org/10.1016/j.dcn.2016.05.004

Yamagishi, T, Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., Miura, A., Inukai, K., Takagishi, H., & Simunović, D. (2012). Rejection of Unfair Offers in the Ultimatum Game Is No Evidence of Strong Reciprocity. *Proceedings of the National Academy of Sciences, 109,* 20364-20368. https://doi.org/10.1073/pnas.1212126109

Yamagishi, T., Li, Y., Fermin, A. S. R., Kanai, R., Takagishi, H., Matsumoto, Y., Kiyonari, T., & Sakagami, M. (2017). Behavioural Differences and Neural Substrates of Altruistic and Spiteful Punishment. *Scientific Reports, 7,* Article No. 14654. https://doi.org/10.1038/s41598-017-15188-w

Zourrig, H., Chebat, J.-C., & Toffoli, R. (2015). "In-Group Love and Out-Group Hate?" A Cross Cultural Study on Customers' Revenge, Avoidance and Forgiveness Behaviors. *Journal of Business Research, 68,* 487-499. https://doi.org/10.1016/j.jbusres.2014.09.014