

J

Justine Effect



Jaroslav Flegr
Department of Philosophy and History of
Science, Faculty of Science, Charles University,
Prague, Czech Republic

Synonyms

None. Related concepts are antisocial punishment and perverse punishment.

Definition

The Justine effect refers to the positive correlation between a player's contribution and the amount of punishment they receive in the Public Goods Game with punishment, observed specifically in approximately the top quartile of the most altruistic players.

Altruistic Punishment in a Public Good Game

The study of economic behavior, psychology, and ethology often utilizes experimental games, with the Public Good Game (PGG) being a popular choice for exploring human altruism and cooperation. In the PGG, individual anonymous players are given a certain amount of money, which they

can choose to contribute, in whole or in part, to a common pool. The total amount in the pool is then doubled and distributed equally among all players, regardless of their individual contributions.

In the variant of PGG, the Public Good Game with Punishment (PGGP), players are given the opportunity to purchase the right to punish other players based on the amount they contribute. Despite the cost of punishing others, players often punish those who contribute little or nothing to the pool, thereby creating a system of altruistic punishment that encourages cooperation and high levels of contribution to the common pool. This type of punishment is purely altruistic, with the cost of punishment being borne solely by the punisher and the potential benefits being shared by all players.

Antisocial Punishment

Delving deeper into Public Goods Game with Punishment reveals a surprising phenomenon – antisocial punishment. This manifests in the “Public Goods Game” as some players punish those who are cooperative and contribute more than sufficiently or even the most, to the group. While punishment is generally considered to sustain cooperation within a group, antisocial punishment can undermine it by discouraging the most altruistic players from continuing to contribute. Research has shown that antisocial punishment

is a prevalent phenomenon in the PGGP, with a significant proportion of punishments being misused in this way. For example, Cinyabuguma et al. (2006) claim that “typically 20% or more” of punishments are misused in this way. A cross-cultural experiment conducted in 16 cities of the world reports a range of 6% (Melbourne) to 48% (Muscat) (Herrmann et al., 2008). This figure is too large for the antisocial punishment to be considered a marginal phenomenon.

Discovery of the Justine Effect

The purpose of antisocial punishment is not entirely clear. While some researchers have suggested that it reflects a desire for revenge or a response to perceived injustice, others have argued that it may be a strategic attempt to maximize one’s own payoff by targeting the most altruistic players. Another hypothesis suggested that the aim of the antisocial punishment is to convince altruistic people to be even more altruistic. If it was true, then the targets of antisocial punishment were the altruistic, but not the most altruistic subjects.

Most published studies have investigated the phenomenon of antisocial punishment in systems with only four players. In these studies, it was not possible to distinguish whether the targets of antisocial punishment were altruistic or the most altruistic players. This question was addressed by a study in 2015 (Kuběna et al., 2014), in which a Public Goods Game with Punishment was played by 12 players. The study showed that the dependence of the player’s punishment on their contribution had a U-shape, i.e., it first decreased with their contribution, but in the last quartile, the dependence reversed, and the probability and amount of punishment increased monotonically. The highest antisocial punishments, therefore, were imposed on the most altruistic players.

The authors of the study named the increase in punishment with the increasing level of proven altruism among altruistic individuals the Justine effect, in honor of the unusually altruistic character of the well-known 1791 novel of Marquis de

Sade (2005). The study also showed that the culprits of antisocial punishment are the most selfish individuals. The authors concluded that malicious players punish their counterparts more when they find them less similar to themselves. As they have no avenue to acquire additional information, the contribution level is the only feature available to them to assess (dis)similarity of a coplayer.

The authors do not discuss the reasons for punishing dissimilar individuals, but it could be suggested that selfish individuals want to retaliate against altruists, whom they (mostly rightfully) consider as people who punished them for their selfishness. However, it is possible that the perpetrators of the Justine effect have economic reasons for their behavior – they reduce the willingness of the least altruistic individuals, who certainly feel punishment as injustice, to punish the selfish ones. This way, the selfish individuals can protect themselves from future punishment.

Conclusions

The Justine effect is an intriguing phenomenon observed in the Public Goods Game with Punishment, in which the most altruistic players are more likely to be punished than their less altruistic counterparts. This effect suggests that altruism can actually be a liability in a cooperative setting, as it makes one more vulnerable to punishment. It is not entirely clear why some players target the most altruistic individuals, but it could be an attempt to reduce the willingness of the least altruistic players to punish the selfish ones. Understanding the Justine effect and its underlying mechanisms may provide insights into dark sides of the human psyche, including the “tall poppy syndrome,” where successful or high-achieving individuals are targeted for criticism or punishment by those who are less successful. In a lighter note, the Justine effect may also provide a new lens through which to view the fate of the eponymous character in the novel *Justine*, whose altruistic actions ultimately led to her downfall.

Cross-References

- ▶ [Altruistic Punishment](#)
- ▶ [Antisocial Punishment](#)
- ▶ [Perverse Punishment](#)
- ▶ [Public Good Game](#)

References

Cinyabuguma, M., Page, T., & Putterman, L. (2006). Can second-order punishment deter perverse punishment?

Experimental Economics, 9(3), 265–279. <https://doi.org/10.1007/s10683-006-9127-z>

Herrmann, B., Thoni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, 319(5868), 1362–1367. <https://doi.org/10.1126/science.1153808>

Kuběna, A. A., Houdek, P., Lindová, J., Příplatová, L., & Flegr, J. (2014). Justine effect: Punishment of the unduly self-sacrificing cooperative individuals. *PLoS One*, 9(3). <https://doi.org/10.1371/journal.pone.0092336>

Marquis de Sade, D. A. F. (2005). *Justine, or, The misfortunes of virtue: A philosophical romance*. Barnes & Noble Books.