Proceedings of the 2nd International Conference on Interdisciplinary Humanities and Communication Studies
DOI: 10.54254/2753-7048/31/20231878

The Impact of Punishment Level and Distribution Ratio on Second Party in Third-Party Punishment Game

Victoria Hope Xu[1,a], Elizabeth Gloria Xu [2,b,*]

[1]WLSA Shanghai Academy, Yang Pu District, Shanghai, 200433, China
[2]Shanghai High School International Division, Xu Hui District, Shanghai, 200231, China
a. victoria_h_xu@163.com, b. may_march25lily@163.com
**corresponding author*

Abstract: The third-party punishment game (TPP) was popular for its enormous influence on real-life judicial problems and the current justice system. After mimicking a third-party game for each participant, the study collected data from the second party and tested the likeableness of the third party and the extent to which the second party punished or rewarded the third party. Most results were calculated by SPSS, and results showed that (i) the second party is more likely to rate a higher level of likeableness and justice for the third party for each level of punishment (ii) if the third party enforces a high level of punishment, the second party is more likely to compensate with a greater amount (iii) if the third party chooses the low level of punishment, the second party is more likely to punish. Overall, the experiment demonstrates how the second parties' attitudes to third parties were influenced by the punishment level, the distribution ratio and third parties' intervention intentions.

Keywords: third party punishment game (TPP), second party, punishment

1. Introduction

Living in such a large society, humans must allocate resources to sustain the group. However, this allocation is not always equitable, as seen in extreme cases such as dictatorships. To examine the psychological response to dictatorship, researchers have attempted to replicate real-life situations through modeling. The dictator game, in which one person in the game has absolute control over the allocation of resources, was first introduced by Daniel Kahneman. He designed the game to involve two to three participants, naming them Player A and Player B. The aim was to determine whether people would be more inclined to distribute the resources evenly, so that both individuals share an equal portion of $20, or to distribute them unevenly. Since then, this study has gained significant popularity and has been tested with various variables. This led to the development of a related game called the third-party punishment game. This third party, often driven by a sense of justice, becomes involved and is granted the power to punish [1]. In this game, individuals belonging to the third party punish violators without any material gain, even though the punishment may cost them their own money. This punishment is often driven by a sense of justice. The initial goal of this game is to test the relationship between punishment and social norms.

Extending from this goal, numerous previous studies have examined whether an increased violation of social norms will result in an increase in punishment from third parties. Ernst Fehr and Urs Fischbacher's study, focusing on an evolutionary basis, clearly shows that as more social norms

DOI: 10.54254/2753-7048/31/20231878

are violated, the level of punishment increases [2]. Furthermore, other studies have compared the results across different human societies and found that people from various societies tend to have a similar preference for increased punishment as unequal behavior increases [3]. In 2012, Marijke C. Leliveld introduced a choice of compensation for the third party. The study aimed to examine how the third party's empathic concern affects their willingness to either punish or compensate [4]. This topic further progressed into the realm of trustworthiness. Jillian J. Jordan concluded that punishers were less inclined to solely punish, opting instead to help in order to signal trustworthiness [5]. The historical study of third-party punishment illustrates a shift in focus over time: researchers initially were interested in studying the initial goal of third-party punishment, and then they dug into the factors that affect the choices of the third party [6]. Yet, the majority of studies have primarily focused on understanding the behavior of the third party, while not paying as much attention to the second party, who lacks the power to allocate resources or administer punishment.

In this paper, the researchers 'exploration contributes to the understanding of the effect of altruistic punishment by exploring the second party's attitude towards the penalties given by the third party. The paper discusses the results under different conditions, specifically by giving different distribution ratios and punishment levels to test the second party's reaction. This can be evaluated by two aspects: one is the level of the second party's likeableness toward the third party, and another is to what extent the second party feel the third party's action is justified. The paper will also give the second party the power to punish after the third party punishes, further testing on second party's attitude.

2. Method

Eighty-six teenagers in Shanghai, China participated in this experiment. In specific, instead of being the one to allocate money, participants are assigned as Player B, the one who merely accepts the results of allocation from Player A and Player C and does not have any decisive power. Before the testing, participants were told that their names and any decisions in the game were fully anonymous and their privacy was fully protected. They were told to randomly select the number 1, 2, or 3. In fact, the participants would be assigned to Player B regardless of the number they chose. They were informed that they would participate in a game with two actual players named Player A and Player C and that they would receive real-time feedback from Player A. Then, the researchers would tell the participant that they recently had to share $100 money with Player A, without mentioning where the money comes from. During the allocation, Player A has absolute power and the participant is not able to refuse any decisions made by Player A. At the same time, Player C will watch the process of allocation. Then, the researchers would pause for about 5 to 10 seconds, mimicking the decision-making process of a real Player A. After that, the participant is told the allocation decision made by Player A, which includes three possibilities: (1) The participant will receive $50, while Player A will receive $50, resulting in an even distribution. (2) The participant will receive $10, while Player A will receive $90, showing an obviously imbalanced distribution. (3) The participant will receive $70, while Player A will receive $30, revealing a slightly imbalanced distribution. After the participant receives the results, the researchers introduce the next stage where Player C gets to decide. A stimulated Player C will determine if they want to enforce punishment on Player A and choose between three choices: no punishment, weak punishment (deducting $3 from Player C and $9 from Player A), or strong punishment (forcing Player A and Player B to have an equal distribution).

3. Study 1

3.1. Procedure

Based on the experiment introduced in the introduction, the researchers add questions based on participants' attitudes towards Player C. To further quantify the results, the researchers asked the

DOI: 10.54254/2753-7048/31/20231878

participants about their likeableness toward Player C rated on a scale of 1-5, with 1 showing a very low likeableness, and 5 showing a very high likeableness. Additionally, the participants are also asked about their feelings of justice for Player C, who is the third party in the experiment. The level of justice is also rated on a scale of 1-5 with the rule exactly the same as the rule for likeableness.

From the paper's perspective, the rating of likeableness symbolizes the general impression, reflecting the trustworthiness of Player C. The paper wants to study if Play B's feeling of trustworthiness of Player C depends on the extent of unjust treatment. Accordingly, the paper hypothesizes when Player B receives more unjust treatment (90:10 > 70:30 > 50:50), they view Player C who punishes the dictator with higher trustworthiness and justice, within the punishment level.

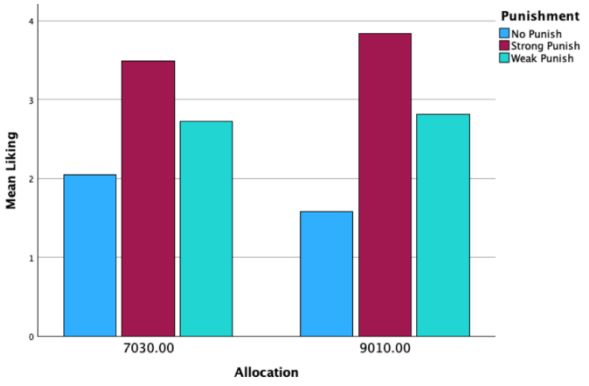

Figure 1: Mean liking for each allocation level and punishment level.

3.2. Result

Considering an unfairness without the participation of the third party, the liking rate from Player B under such a condition may reflect the authentic feeling regardless of any punishments reinforced by the third party. Thus, the liking rate with no punishment will be considered as the control group for each condition.

The bar graph (Fig. 1) demonstrates the mean liking rate from Player B towards the artificial Player C under 50:50, 70:30, and 90:10 allocation. For statistical convenience, the researchers select the mean value of the liking rate as a representative of the group (n=86)'s liking rate. Under the allocation of 70:30 and strong punishment from Player C, the mean liking rate is 3.49; while in the same case but under weak punishment, the mean liking rate drops to 2.72. Similarly, under the allocation of 90:10 and strong punishment from Player C, the mean liking rate is 3.84; while in the same case but under weak punishment, the mean liking rate decreases to 2.81.

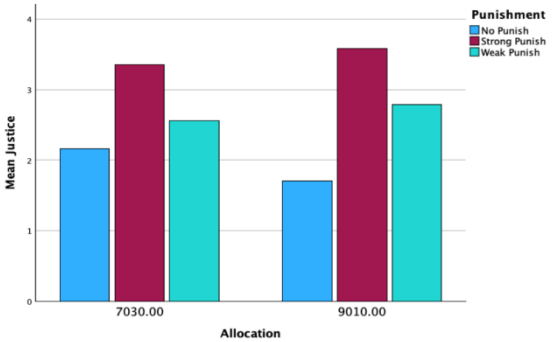

Figure 2: Mean justice for each allocation level and punishment level.

DOI: 10.54254/2753-7048/31/20231878

For the second bar graph (Fig. 2), the "No Punish" group also serves as the control group for a baseline of comparison. As the graph shows, within the 86 participants tested, while the mean level of justice the participant rated for the normal group is only 2.1628 for the distribution level of 70:30, the mean reaches 2.5581 when the third party enforces weak punishment. Moreover, when the punishment reaches a higher level, the participant rated the third party as justified with a mean level of 3.3488 which is more than 2.5 (neutral). For a distribution level of 90:10, however, the control group reaches a lower mean of justice of 1.6977, which indicates a sign of not justice. Comparatively, when the third party enforces a weak punishment or strong punishment, both means of justice rated by the second party are higher than that of the 70:30 distribution level. This indicates that within a more unbalanced distribution level, the second party tends to feel more justice for the third party if the same level of punishment level is enforced.

These data align with the initial hypothesis, as they demonstrate that when Player B experiences more extreme unfairness, the liking rate and level of justice rated for the third-party increase, regardless of the level of punishment. However, it is important to note that the participants' liking is largely influenced by the actions taken by Player 3, rather than the specific circumstances they find themselves in. This conclusion coincides with Singer et al.'s investigation of victim perception. They found out that victim's perception of justice and liking is positively correlated with third-party punishment [7]. Consequently, participants may perceive the individual who contends for their benefits the most as more reliable than others. The results further illustrate a correlation between the liking rate and the level of justice rated. As the liking rate for the third-party increases, the sense of justice towards the third party also increases, as evidenced by the highest levels of likeableness and perceived justice towards the third party observed for strong punishment within each distribution ratio.

4. Study 2

In addition to examining the impressions of trustworthiness and justice, the paper also aims to investigate the actions that Player B will take towards the helper, Player C. Unlike rating one's feelings, actions reflect deeper concerns of Player B that may not be solely explained by considerations of justice or trustworthiness, but also by their perception of Player C's role and obligation. These actions will be evaluated through another form of punishment administered by Player B. In line with this, previous studies conducted by Hechler et Kessler have also explored the punishment administered by Player B, the victim, with the intention of comparing the level of punishment implemented by a sufferer and a bystander. [8] They ultimately figured out that the sufferer tends to administer more severe punishment compared to a bystander in response to an unfair condition. Furthermore, the studies presented by Kanakogi Y emphasize that any interventions or actions taken by the third party can mitigate the victims' inclination towards aggression and extreme behavior [9]. This is because the victims may perceive that the dictators have undergone a value transformation due to the punishment imposed by the third party [10].

Although the previous study focused on the punishment administered by the sufferer to the bystander, it does offer valuable insights. As a result, the experimenters in this study hypothesized that the actions taken by Player B would depend on the outcome of the allocation after punishment. If Player B perceives the outcome as just, it is expected that they will reward Player C moderately, and penalize Player C more severely if they neglect the unjust situation.

4.1. Procedure

The paper established nearly the exact same third punishment game in the study, despite that Player B is not asked about the rating of Player C. Instead, after Player C made a decision based on the

DOI: 10.54254/2753-7048/31/20231878

distribution ratio Player A decided, Player B was informed to have decisive power. Player B can now decide whether to punish Player C or reward Player C without letting Player C know. The researchers will first ask whether participants want to use this decisive power. If the participant responds yes, then the researchers will ask whether to punish or reward and for how many dollars. Researchers will also notice that rewarding or punishing will not take away anyone's money. If the participant decides to reward, Player C's money will simply be added. If the participant decides to punish, Player C's money will just be deduced at the exact same amount the participant announced. Rewarding will result in any number with a positive sign. Say if the participant decides to reward Player C $20, then the data will show +20. If the participants are punished for $40, then the data will be written as -40. After the whole process, the researchers will then question the participant's mood and feelings, further investigating the reason why the participant makes that decision.

Based on the previous rule, the paper records the mean number of compensations the second party gives to the third party for each punishment level. Within each punishment level, the mean number of compensations in each distribution level is compared. As shown by the box part, 0 means an average of no punishment or compensation. Any number below 0 represents that average more punishments are enforced. Any number above 0 represents that average more compensation is provided. The number within the box represents 75% of the data collected for each distribution ratio and punishment ratio. Star point represents outliers that are 1.5IQR below 25% or above 75%. From that, the paper finds out the relationship between compensation or punishment with distribution ratio and punishment level.

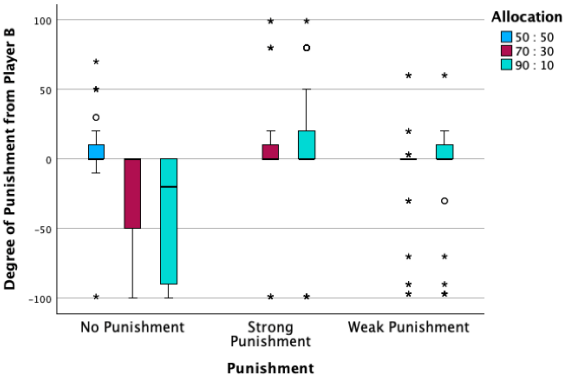

Figure 3: Degree of player B's punishment towards player C under various allocation and third-party punishment.

4.2. Results

The box plot (Fig. 3) shows that when the third party enforces no punishment if the distribution ratio by Player A is 50:50, 75% of the participants will choose to grant extra compensation. However, if the distribution ratio is uneven-either by 70:30 or 90:10- will mostly result in punishment. For 70:30 distribution, if the third party choose not to punish, the third party will get between 0 and $50 punishment. If the distribution ratio is more extreme, that is to say, it reaches 90:10, most of the participants choose to punish in a range of 0 to $80 punishment, with a mean of $37.35 punishment. However, for the same distribution level, as the third party enforces weak punishment, the average number significantly increases. For the 70:30 distribution level, the third party received an average of $12.37 punishment for weak punishment, while a $8.16 compensation is provided within a strong punishment. The result is slightly similar for the 90:10 distribution level, where the third party received $8.12 punishment with weak punishment and $7.98 compensation with strong punishment.

DOI: 10.54254/2753-7048/31/20231878

The results show that as long as there is uneven distribution, no punishment will always lead to punishment from the second party. As the distribution ratio is more uneven, the punishment from the second party increases, coinciding with the hypothesis. However, if the distribution is even, even if the third party does nothing, the second party still compensates for the third party.

In the case of uneven distribution, all participants choose to punish, indicating that a bystander will indeed be treated unfavorably by the victim. As anticipated, the severity of the punishment is positively correlated with the level of uneven allocation. However, in situations where the third party imposes additional punishment on the dictator, participants typically respond with neutral or positive actions. This may be explained by the tendency of individuals to invest in providing positive feedback [11]. Or it can be explained by the neural investigation conducted by Fehr, suggesting humans are evolutionary rewarding altruistic behaviors [12]. The compensation provided by the second party, thus, may be regarded as a reward for the altruistic behaviors made by the third party.

5. Discussion

This study provides valuable insights into real-life situations, enabling the exploration of ideas related to justice. On a smaller scale, people can apply these findings to judge and assess daily problems that involve a third-party punisher. However, the application of this study extends to a much larger scale. For instance, it can be utilized in cases where a worker receives unfair payment from their employer. Even if the employer faces punishment from a union, the worker may still not feel satisfied. Similarly, in criminal cases, although the justice system imposes punishment on the offender, the victim may not feel justified. This study aids in evaluating and predicting the reactions of the second party, thereby contributing to a more sophisticated understanding of the concept of justice.

Some of the limitations of this study include the limited ways of intervening in the third-party punishment game. In real-life cases, a third party can also choose to compensate the second party instead of punishing the first party. Setting the third party's choice as weak or strong punishment toward the first party may lead to limited application in real life. Therefore, the future study can further explore the second party's attitudes toward the third party after the third party compensate the second party.

6. Conclusion

Overall, the paper explores the second party's attitudes from the perspectives of internal attitude and external behavior, which not only proves victims' expectations and demands for third-party intervention in unfair events but also finds that this effect is regulated by the degree of first-party unfairness and the intensity of third-party punishment. The results from Study 1 show that the rating of likeableness and justice for third parties are also highly correlated, showing that as the rating of likeableness increases, the rating of justice also increases. Results from Study 2 reveal that for uneven distribution if the third party enforces a high level of punishment, the second party is more likely to compensate with a greater amount. Inversely, if the third party chooses a low level of punishment, the second party is more likely to punish with a greater amount.

References

[1] *Lotz, S., Baumert, A., Schlösser, T., Gresser, F., & Fetchenhauer, D. (2011). Individual differences in third-party interventions: How justice sensitivity shapes altruistic punishment. Negotiation and Conflict Management Research, 4(4), 297-313.*

[2] *Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. Evolution and human behavior, 25(2), 63-87.*

[3] *Engel, C. (2011). Dictator games: A meta study. Experimental economics, 14, 583-610.*

DOI: 10.54254/2753-7048/31/20231878

[4] *Leliveld, M. C., van Dijk, E., & van Beest, I. (2012). Punishing and compensating others at ythe own expense: The role of empathic concern on reactions to distributive injustice. European Jthenal of Social Psychology, 42(2), 135-140.*

[5] *Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. Nature, 530(7591), 473-476.*

[6] *Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., ... & Ziker, J. (2006). Costly punishment across human societies. Science, 312(5781), 1767-1770.*

[7] *Singer, T. (2006, October). The neuronal basis of empathy and fairness. In Empathy and Fairness: Novartis Foundation Symposium 278 (pp. 20-40). Chichester, UK: John Wiley & Sons, Ltd.*

[8] *Hechler, S., & Kessler, T. (2022). The importance of unfair intentions and outcome inequality for punishment by third parties and victims. Zeitschrift für Psychologie.*

[9] *Kanakogi, Y., Inoue, Y., Matsuda, G., Butler, D., Hiraki, K., & Myowa-Yamakoshi, M. (2017). Preverbal infants affirm third-party interventions that protect victims from aggressors. Nature Human Behavithe, 1(2), 0037.*

[10] *Gromet, D. M., Okimoto, T. G., Wenzel, M., & Darley, J. M. (2012). A victim-centered approach to justice? Victim satisfaction effects on third-party punishments. Law and Human Behavior, 36(5), 375.*

[11] *Raihani, N. J., & Bshary, R. (2015). Third-party punishers are rewarded, but third-party helpers even more so. Evolution, 69(4), 993-1003.*

[12] *Fehr, E., & Rockenbach, B. (2004). Human altruism: economic, neural, and evolutionary perspectives. Current opinion in neurobiology, 14(6), 784-790.*