The Effect of Belief in a Just World on Third-Party Punishment: A Three-Level Meta-Analysis

Authors: Huang Chuanbin, Wang Yang, Ding Yi, Guo Yongyu, Guo Yongyu

Date: 2026-01-26T18:10:55+00:00

Abstract

When witnessing injustice, should one draw the sword to help? Whether belief in a just world facilitates third-party punishment remains a matter of debate in both theoretical hypotheses and empirical research. To clarify this controversy, the present study adopts a three-level meta-analytic approach to systematically examine the relationship between belief in a just world and third-party punishment, as well as its moderating factors. Through literature search and screening, 46 articles and 93 effect sizes were obtained, comprising a total of 15,772 participants. The results revealed a significant but weak positive correlation between belief in a just world and third-party punishment (r = 0.08, 95% CI = [0.03, 0.14]). This relationship was moderated by the self-other dimension of belief in a just world, the form of third-party punishment behavior (social punishment vs. monetary punishment), cultural background, type of norm violation, and sample type, but not by gender, age group, the explicit–implicit dimension of belief in a just world, the type of third-party punishment (punitive behavior vs. willingness to punish), or research method. From the perspective of third-party punishment, this study systematically tests the competing hypotheses of just-world belief theory and their boundary conditions, providing implications for subsequent theoretical development and empirical research.

Full Text

The Effects of Belief in a Just World on Third-Party Punishment: A Three-Level Meta-Analysis

HUANG Chuanbin, WANG Yang, DING Yi, GUO Yongyu
(School of Psychology, Nanjing Normal University, Nanjing 210097, China)

Abstract

Does seeing injustice compel one to draw their sword against it? Whether belief in a just world promotes third-party punishment remains theoretically and empirically contested. To clarify this controversy, this study employed a three-level meta-analysis to systematically examine the relationship between belief in a just world (BJW) and third-party punishment, along with potential moderating factors. Through comprehensive literature search and screening, we obtained 46 studies comprising 93 effect sizes with a total of 15,772 participants. Results revealed a significant yet weak positive correlation between belief in a just world and third-party punishment ($r = 0.08$, 95% CI = [0.03, 0.14]). This relationship was moderated by the self-other dimension of BJW, the form of third-party punishment behavior (social vs. monetary punishment), cultural background, violation scenario, and sample type, but not by gender, age group, the explicit-implicit dimension of BJW, type of third-party punishment (punishment behavior vs. punishment intention), or research methodology. This study systematically tests competing hypotheses of just world theory from a third-party punishment perspective and specifies their boundary conditions, providing insights for future theoretical development and empirical research.

Keywords: belief in a just world, third-party punishment, three-level meta-analysis, moderating effect

1. Introduction

The pursuit of justice is fundamental to human nature (Guo & Hu, 2015; Chomsky & Foucault, 2011). Throughout social development, humans have gradually formed justice-oriented social norms that constrain behavior for oneself and others. As a crucial mechanism for maintaining such norms, third-party punishment refers to the punitive intentions and behaviors toward transgressors exhibited by uninvolved third parties who have no direct stake in the violation (Fehr & Fischbacher, 2004; Fehr & Gächter, 2002; Molenmaker et al., 2016). Third-party punishment not only maintains social norms by restoring balance and promoting norm compliance (Guo et al., 2024), but also plays a vital role in fostering social fairness and interpersonal cooperation (Chen et al., 2021; Martin et al., 2021; McAuliffe et al., 2025). Yet in real life, why do some witnesses of injustice choose to "see injustice and draw their sword to help," while others remain indifferent or even kick someone when they' re down?

Since its proposal, just world theory has focused on this phenomenon (Lerner, 1980; Lerner & Simmons, 1966), but the theory contains multiple competing hypotheses that limit its explanatory power (Hafer & Bègue, 2005; Hafer & Rubel, 2015; Lerner, 1980). These theoretical disagreements are particularly prominent in empirical research on belief in a just world and third-party punishment, yielding three contradictory conclusions: facilitative effects (Mikani & Rafiee, 2023; Lin et al., 2024), inhibitory effects (Chahal et al., 2022; Huang et al., 2024), and null effects (Ashdown et al., 2024; Shechory-Bitton & Zvi, 2020). To clarify

these competing hypotheses within just world theory and resolve controversies in third-party punishment research, this study employs three-level meta-analysis to quantitatively integrate and analyze the relationship between belief in a just world and third-party punishment, as well as factors that may influence this relationship, thereby providing insights for future theoretical development and empirical exploration.

1.1 Theoretical Hypotheses and Empirical Evidence on BJW's Effects on Third-Party Punishment

Belief in a just world (BJW) refers to the conviction that the world is a just place where people get what they deserve and deserve what they get (Zhou & Guo, 2013). However, injustice is ubiquitous in real life. According to cognitive dissonance theory, when individuals encounter information inconsistent with their BJW, they experience cognitive dissonance (Festinger, 1957). To reduce this discomfort and restore a sense of justice, just world theory proposes three coping strategy hypotheses (Ellard et al., 2016; Hafer & Rubel, 2015; Lerner, 1980), all of which have received empirical support.

Among these, the irrational strategies hypothesis of BJW has received the most attention. It suggests that people tend to rationalize injustice through cognitive reconstruction to restore justice at the cognitive level (Zhou & Guo, 2013). This coping strategy posits that BJW typically leads to victim blaming (Hafer & Rubel, 2015; Lerner, 1980) and inhibits third-party punishment (Chahal et al., 2022; Valor-Segura et al., 2011). For example, a U.S. online survey found that after reading about school bullying, individuals with high BJW showed less sympathy for victims, were more likely to blame victims, experienced less anger toward bullies, and tended to exonerate bullies regardless of whether participants were reminded that the witnessed event constituted bullying (Voss & Newman, 2021). Huang et al. (2024) also found through questionnaire methods that after reading about celebrity infidelity scandals, high BJW individuals exhibited fewer negative emotions toward unfaithful behavior and consequently reduced blame for celebrity infidelity. Another study with Swedish citizens provided evidence for the antisocial function of BJW, finding that while the public generally preferred punishing transgressors over blaming and derogating victims to restore justice, high BJW individuals tended to blame sexual violence victims rather than perpetrators (Strömwall et al., 2012). In summary, nearly 60 years of accumulated literature seems to collectively indicate that belief in a just world is a "fundamental delusion" (Dawtry et al., 2020; Lerner, 1980; Rawlings et al., 2024), or even an antisocial or defensive belief system (Furnham, 2003).

However, the rational strategies hypothesis of BJW challenges this antisocial view (Bal & van den Bos, 2015), proposing that BJW, as a justice motive, can promote positive justice behaviors to restore justice in reality (Zhou & Guo, 2013), such as preventing and punishing transgressors (Dalbert, 2001; Lerner, 1980; Ellard et al., 2016). This view is supported by the deontic model of justice, which suggests that when witnessing injustice, observers' justice motives drive

them to punish transgressors based on moral intuition to restore justice norms (Folger, 2001). Empirical research also shows that BJW can motivate people to actively pursue and maintain justice. For instance, a study with Chinese university students found that after priming BJW through textual materials, the high BJW activation group exhibited more punishment behavior in third-party game paradigms and overestimated others' punishment levels (Yang & Chen, 2022). Previous meta-analyses have found that BJW is significantly negatively correlated with antisociality (Kong et al., 2021) and positively correlated with prosocial personality traits such as extraversion and agreeableness (Nudelman, 2013). Additionally, indirect reciprocity theory provides a basis for the prosocial function of BJW (Nowak, 2006): third-party punishment and other altruistic behaviors typically involve resource costs, but BJW emphasizes reciprocal norms of "good begets good, evil begets evil," prompting people to focus on indirect benefits such as reputation enhancement and future cooperation (Jordan et al., 2016; Xiao & Wang, 2025). This positive expectation of future outcomes helps individuals better tolerate short-term resource losses (Li et al., 2022), thereby strengthening motivation to implement third-party punishment. However, only limited indirect evidence supports this mechanism. For example, one questionnaire study found that BJW could increase prosocial behavior by enhancing psychological resilience (Liu et al., 2025).

Furthermore, the protective strategies hypothesis of BJW has also received empirical support. Specifically, high BJW individuals may ignore and avoid injustice threats by emphasizing the existence of ultimate justice or adopting a "multiple-world view" (Ellard et al., 2016; Hafer & Rubel, 2015), resulting in no real-life action (Zhou & Guo, 2013). In other words, BJW and third-party punishment may be unrelated. Previous studies have also found bystander effects in unjust situations. For example, in a study examining how individuals respond to charitable appeals, participants were asked to walk alone along a predetermined route where they would pass a table displaying different help-seeking information. Results showed that compared to situations presenting only charitable donation slogans or no help-seeking information, when high-salience help-seeking content (e.g., photos of impoverished children) was displayed, participants maintained greater physical distance from the table. More importantly, when facing high-salience help-seeking information, high BJW individuals showed significantly poorer memory for details of the appeal compared to low BJW individuals (Pancer, 1988). This suggests that high BJW individuals may actively avoid high justice-threat information through dual cognitive and behavioral avoidance strategies to maintain their BJW. Similarly, other studies have found that after reading about sexual violence, high BJW individuals neither rationalized the status quo by blaming victims (Pedersen & Strömwall, 2013) nor restored justice norms by punishing transgressors (Shechory-Bitton & Zvi, 2020).

In summary, regarding the relationship between belief in a just world and third-party punishment, the three coping strategy hypotheses within just world theory make different predictions, all of which have received empirical support. The

irrational strategies hypothesis suggests that BJW, as a system-justifying belief, inhibits third-party punishment. The rational strategies hypothesis proposes that BJW, as a justice motive, facilitates third-party punishment. The protective strategies hypothesis suggests that BJW and third-party punishment may be unrelated. To further examine their relationship and clarify competing hypotheses within just world theory, this study explores possible boundary conditions.

1.2 Moderating Variables in the BJW-Third-Party Punishment Relationship

Inconsistent results regarding BJW' s effect on third-party punishment may be related to study variables (BJW type, third-party punishment type, punishment behavior form), contextual characteristics (cultural background, violation scenario), demographic features of participants (gender, age group, sample type), and research methodology.

1.2.1 Types of Belief in a Just World The dual effects of BJW on third-party punishment may stem from its structural differences. BJW is typically distinguished as BJW-self (belief that the world is just for oneself) and BJW-other (belief that the world is just for others) (Zhou & Guo, 2013). Although positively correlated (Sutton et al., 2017), BJW-self is typically associated with prosocial behavior, whereas BJW-other relates to callous social attitudes (Bartholomaeus & Strelan, 2019). Numerous studies have found that BJW-self not only generates sympathy and compensation behavior toward victims (Chen et al., 2023; Silver et al., 2015) but also encourages punishment of transgressors (Ge, 2014; Mikani et al., 2023; Modesto et al., 2020). In contrast, BJW-other leads people to blame victims for their misfortune (Mikani & Rafiee, 2023; Witte & Flechsenhar, 2025), reducing the likelihood of punishing transgressors (Adolfsson & Strömwall, 2017; Chahal et al., 2022; Huang et al., 2024). Therefore, the self-other dimension of BJW may differentially influence third-party punishment.

The dissociation model of the justice motive attempts to explain BJW' s dual functions from a dual-process perspective (Dalbert, 2012; Kahneman, 2011). Based on this, previous research has distinguished explicit and implicit BJW (Dalbert, 2012). The former refers to conscious descriptions of justice values that produce socially normative "deliberative" responses, while the latter refers to subconscious justice knowledge that generates automatic moral intuition responses (Dalbert, 2012). Consequently, implicit BJW may produce stronger tendencies to punish transgressors than explicit BJW. Previous research has found that implicit BJW shows more stable relationships with third-party punishment. For example, after priming implicit BJW through Implicit Association Test (IAT) paradigms, high implicit BJW individuals in third-party game paradigms were willing to sacrifice personal interests to punish unfair distributions regardless of proposal fairness (Hu, 2024; Yun, 2020). However, high explicit BJW individuals only showed stronger third-party punishment tendencies

under high-unfairness conditions (Yun, 2020). Other studies have found complex relationships between explicit BJW and third-party punishment (Adolfsson et al., 2017; Chahal et al., 2022; Huang et al., 2024). Thus, the explicit-implicit dimension of BJW may also differentially affect third-party punishment.

1.2.2 Types of Third-Party Punishment The relationship between BJW and third-party punishment may also be influenced by punishment type. Third-party punishment refers to punitive intentions and behaviors toward transgressors by uninvolved third parties. Punishment intention represents the subjective willingness to punish, including responsibility attribution, punitive attitudes, and punishment expectations (Hammond et al., 2011; Mikani et al., 2023; Modesto et al., 2020). Punishment behavior refers to actual punitive actions, including monetary punishment and social punishment (Cui et al., 2017; Shen et al., 2025). According to the social intuitionist model, punishment involves a sequential processing stage (Haidt, 2001). First, unjust situations activate automatic BJW-based responses, generating punishment intentions (Lerner, 2003). When the cognitive system engages in cost-benefit weighing, self-interested rationality may inhibit punishment implementation (Yin et al., 2019; Zheng et al., 2024). Therefore, BJW may differentially affect punishment intentions versus behaviors. Previous research shows that the relationship between BJW and punishment intention is relatively stable, with high BJW individuals typically showing more negative attitudes toward transgressors (Voss & Newman, 2021), greater belief that transgressors should be held responsible (Hammond et al., 2011), and expectations that they should be punished (Modesto et al., 2020). However, the relationship between BJW and punishment behavior is more complex (Yang & Chen, 2022; Russell & Hand, 2017; Shechory-Bitton & Zvi, 2020). Thus, to test whether BJW differentially affects punishment intentions and behaviors, this study examines the moderating role of third-party punishment type (punishment behavior vs. punishment intention).

1.2.3 Forms of Third-Party Punishment Behavior The BJW-third-party punishment relationship may also be influenced by punishment behavior form. Third-party punishment behaviors include monetary punishment and social punishment (Cui et al., 2017; Shen et al., 2025). Monetary punishment involves reducing transgressors' monetary gains, while social punishment involves non-material forms such as verbal criticism (Cui et al., 2017). Previous research using third-party game paradigms has predominantly found that high BJW individuals tend to use monetary punishment (Chen et al., 2022; Yang & Chen, 2022). However, in studies using social scenarios, the relationship between BJW and social punishment is more complex, with some studies finding negative correlations (Huang et al., 2024; Russell & Hand, 2017) and others finding positive or null relationships (Adolfsson et al., 2017; Levy & Reuven, 2018). Therefore, the form of third-party punishment behavior (social vs. monetary punishment) may also moderate the BJW-third-party punishment relationship.

1.2.4 Cultural Background The effect of BJW on third-party punishment may vary across cultures. Although myths and nursery rhymes about "good begets good, evil begets evil" exist across cultures (Rubin & Peplau, 1975; Lucas et al., 2016), previous BJW research shows clear cultural bias. Specifically, individualistic cultures emphasize personal rights and well-being (Kitayama & Salvador, 2024), typically basing moral agency on an individualized self unrestricted by social interdependence, causing justice sensitivity to become subordinate to personal interest considerations (Wu et al., 2025). Many studies have found that individuals in individualistic cultures are primarily motivated by self-interest (Markus & Kitayama, 2010), tending to blame victims to protect their own BJW from threat (Lerner, 1980; Dawtry et al., 2020) rather than compensating victims and punishing transgressors (Adolfsson & Strömwall, 2017; Shechory-Bitton & Zvi, 2020). However, in collectivist cultures emphasizing relational interdependence, people value social responsibility and reciprocal morality among group members more highly (Miller et al., 2014; Xiao & Wang, 2025), with justice understanding based on public relational models and principles of responsibility (Brockner et al., 2005; Wu et al., 2025). For example, Wu et al. (2025) found through surveys, priming experiments, and large language model role-playing studies that collectivism enhances prosocial justice sensitivity by fostering shared responsibility. Other researchers have found that in collectivist cultural contexts, high BJW individuals both actively punish transgressors (He et al., 2021) and overestimate the likelihood that others will punish transgressors (Yang & Chen, 2022; Bai et al., 2014). Thus, cultural background may lead to different effects of BJW on third-party punishment.

1.2.5 Violation Scenarios The effect of BJW on third-party punishment may vary by scenario. Violation scenarios refer to specific contexts in which individuals violate social norms. Researchers have examined high BJW individuals' reactions to injustice in scenarios including sexual violence (Adolfsson & Strömwall, 2017), car accidents (Lin et al., 2024), robbery (Mikani & Rafiee, 2023), corruption (Mikani et al., 2023), school bullying (Voss & Newman, 2021), and disease infection (Stinnett et al., 2024). Because these scenarios differ in severity and threat to justice, the coping strategies generated by BJW also differ (Dawtry et al., 2020; Witte & Flechsenhar, 2025). Studies have found that in corruption scenarios challenging procedural justice, both Chinese and Iranian individuals with high BJW reported higher punishment likelihood (Bai et al., 2014; Mikani et al., 2023). However, in sexual violence scenarios threatening personal safety, high BJW individuals tended to blame victims rather than punish transgressors (Ashdown et al., 2024; Russell & Hand, 2017). This suggests that violation scenarios may moderate the effect of BJW on third-party punishment.

1.2.6 Gender The effect of BJW on third-party punishment may differ by gender. According to Gilligan' s (1982) theory of moral development, women tend to focus on others' suffering and maintaining interpersonal relationships, showing a care orientation, whereas men emphasize abstract moral principles

such as fairness and rights, showing a justice orientation (Jaffe & Hyde, 2000). This suggests women are more likely to sympathize with and help victims, while men are more likely to punish transgressors. Previous research has also found that men are more willing than women to pay costs to punish unfairness (Kromer & Bahçekapili, 2010). However, other studies have found that although men have higher BJW levels (Bègue & Bastounis, 2003), they are more likely than women to blame victims (Witte & Flechsenhar, 2025; Ash & Yoon, 2020). Additionally, many studies have found no gender differences in the BJW-third-party punishment relationship (Adolfsson & Strömwall, 2017; Strömwall et al., 2012; Strömwall et al., 2013). Based on these inconsistent findings, this study further examines the moderating role of gender in the BJW-third-party punishment relationship.

1.2.7 Age Groups and Sample Types The effect of BJW on third-party punishment may vary by age or population group. Third-party punishment may increase with children' s age (Yudkin et al., 2019; Smith et al., 2013). Consistently, as cognitive maturity and life experience accumulate, people' s reliance on BJW to avoid meaninglessness and despair gradually strengthens (Bartholomaeus & Strelan, 2019; Dalbert & Stoeber, 2006; Sun et al., 2024). For example, Adolfsson and Strömwall (2017) found that after reading about sexual violence, adults were more likely than adolescents to blame transgressors. However, other researchers argue that as abstract logical thinking matures and negative life experiences beyond personal control increase, people become more aware of systemic injustice, which weakens BJW (Tian et al., 2019; Oppenheimer, 2005). For instance, a meta-analysis found that the negative correlation between BJW and antisociality was stronger in adolescent samples than in university student and non-student samples (Kong et al., 2021). Therefore, to test whether developmental trends exist in the BJW-third-party punishment relationship, this study examines the moderating roles of age group and sample type.

1.2.8 Research Methodology The effect of BJW on third-party punishment may differ by research method. BJW theory has developed two research traditions: experimental and questionnaire studies (Bartholomaeus & Strelan, 2019). The former typically finds that BJW facilitates third-party punishment. For example, previous studies priming BJW through textual materials or recall methods found that high BJW groups were more likely to punish transgressors (Yang & Chen, 2022; Bai et al., 2014). Questionnaire studies, however, reveal more complex relationships, with some finding positive correlations (Lin et al., 2024; Modesto et al., 2020), others negative correlations (Chahal et al., 2022; Huang et al., 2024), and still others null relationships (Ashdown et al., 2024). Therefore, to test whether research methodology influences the BJW-third-party punishment relationship, this study examines its moderating role.

1.3 Research Objectives

In summary, the effect of belief in a just world on third-party punishment shows theoretical and empirical divergence. No study has yet clarified this from an integrative perspective, making it necessary to quantitatively determine the direction, strength, and potential moderators of BJW' s effect on third-party punishment through meta-analysis. Moreover, given that three-level meta-analysis maximally utilizes effect sizes from original literature compared to traditional methods (Assink & Wibbelink, 2016), this study employs three-level meta-analysis to comprehensively examine BJW' s effect on third-party punishment and test whether their relationship differs across BJW type, third-party punishment type, punishment behavior form, cultural background, violation scenario, gender, age group, sample type, and research methodology.

2. Methods

This study followed the Preferred Reporting Items for Systematic Review and Meta-analysis (PRISMA) guidelines (Page et al., 2021). After completing literature screening, main effect analysis, and heterogeneity testing, the study was preregistered on the Open Science Framework (OSF; Registration ID: 10.17605/OSF.IO/MRZAU). To ensure transparency and reproducibility, all materials have been made publicly available on the open science platform (https://osf.io/9usgb).

2.1 Literature Search and Screening

We searched existing domestic and international literature on belief in a just world and third-party punishment. First, in three Chinese databases—CNKI, Wanfang, and VIP—we searched for articles with BJW and third-party punishment in titles or abstracts. BJW search terms included "公正世界信念" (belief in a just world), "公正世界" (just world), "正义信念" (justice belief), and "公正信念" (justice belief). Third-party punishment search terms included "责备" (blame), "谴责" (condemn), "惩罚" (punishment), "第三方利他" (third-party altruism), "第三方惩罚" (third-party punishment), and "第三方干预" (third-party intervention). Second, we searched four English databases: Web of Science, ProQuest, Science Direct, and Google Scholar. BJW search terms included "belief in a just world," "just world," and "justice belief." Third-party punishment search terms included "blame," "derogate," "punishment," "third-party punishment," "third-party altruistic," and "third-party intervention." We combined both keyword categories pairwise for joint searches. Additionally, we used backward citation searching in reviews and included studies to identify additional relevant literature. The most recent update was in September 2025, yielding 1,614 relevant articles.

We imported literature into EndNote X9 and screened according to the following criteria: (1) English or Chinese language only; (2) Empirical studies focusing on the BJW-third-party punishment relationship, excluding theoreti-

cal reviews, case studies, and qualitative research; (3) Studies examining both BJW and third-party punishment, excluding those with only one variable; (4) Articles must report means, standard deviations, sample sizes, and correlation coefficients (r) or convertible statistics (F, t, β). If not reported, we contacted authors; studies were excluded if data could not be obtained; (5) For duplicate publications using the same data, only the most comprehensive report was selected; (6) Participants must be third-party punishers (bystanders), excluding victims or transgressors. The literature screening flowchart is shown in Figure 1 [Figure 1: see original paper].

2.2 Literature Coding and Quality Assessment

First, two authors independently coded each study according to a coding manual (see Appendix C) based on: (A) publication information (author and year); (B) sample size; (C) sample type (middle school/high school/university/non-student/unspecified); (D) mean age and age group (adolescent/early adult/middle adult); (E) female proportion; (F) BJW type (BJW/BJW-self/BJW-other/implicit BJW/implicit BJW-self/implicit BJW-other/unspecified); (G) third-party punishment type (punishment behavior/punishment intention); (H) punishment behavior form (social punishment/monetary punishment); (I) violation scenario type (sexual violence/robbery/corruption/car accident/assault/infidelity/school bullying/labor exploitation/domestic violence/unspecified); (J) cultural background (individualistic/collectivistic/unspecified); (K) research methodology (questionnaire/experiment); and (L) effect size (correlation coefficient). Variable definitions and detailed coding procedures are in the coding manual (Appendix C). Coding followed these principles: (1) Effect sizes were coded by independent sample; multiple independent samples in one article were coded separately; (2) If multiple variables were measured, each was coded separately; (3) Duplicate publications using the same data were coded only once using the most informative source; (4) For longitudinal studies, only the first time point was coded. Final coding was completed independently by two coders. Intercoder reliability was high: ICCs for continuous variables ranged from 0.97 to 1.00, and Kappas for categorical variables ranged from 0.77 to 1.00. Discrepancies were resolved through discussion and consultation of original sources.

Second, we assessed each included study' s quality using the National Institutes of Health (NIH) Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies, scoring each criterion as met (1 point) or not met (0 point) (Lin et al., 2025). Quality ratings were: good (>7 points), fair (5–7 points), and poor (<5 points). Two coders independently assessed quality with Kappa = 0.93. Disagreements were resolved through discussion and consultation of original sources.

2.3 Effect Size Calculation

We used correlation coefficient r as the effect size for the BJW-third-party punishment relationship. When studies reported only F values from one-way ANOVA, t values from independent samples tests, or β values from simple linear regression, we converted these to r values using appropriate formulas (Peterson & Brown, 2005). To control for effects of sample size and normal distribution, we transformed correlation coefficients to Fisher' s z scores for main effect and moderator analyses, then converted back to r for interpretation (Cooper et al., 2019). Effect size magnitude criteria for correlation coefficients were: small ($0.10 \leq r < 0.20$), medium ($0.20 \leq r < 0.30$), and large ($r \geq 0.30$) (Gignac & Szodorai, 2016).

2.4 Model Selection

Most original studies included in this meta-analysis reported multiple effect sizes from the same sample, which are therefore correlated (Cheung, 2014). Traditional meta-analysis ignores this correlation, potentially overestimating overall effect sizes. To overcome dependency issues, many researchers now use three-level analysis to extract multiple effect sizes from the same report, maximizing information retention and statistical power (Lin et al., 2025; Cheung, 2019). Therefore, this study used three-level random-effects models for publication bias assessment, main effect testing, heterogeneity testing, and moderator analysis.

2.5 Publication Bias Control and Assessment

Publication bias refers to the phenomenon where significant results are more likely to be published, causing published literature to inadequately represent the complete body of completed research (Rodgers & Pustejovsky, 2021; Rothstein et al., 2005). To control publication bias, we included both published journal articles and unpublished dissertations. We also used funnel plots, Egger' s regression, and trim-and-fill methods for qualitative and quantitative assessment. Symmetrical funnel plots indicate minimal publication bias (Rothstein et al., 2005). Nonsignificant Egger' s regression results indicate minimal publication bias (Rodgers & Pustejovsky, 2021). When Egger' s regression was significant ($p < 0.05$) or funnel plots showed asymmetrical effect size distribution, we used trim-and-fill to assess publication bias impact. If trimmed-and-filled effect sizes did not change substantially, we concluded that results were minimally affected by publication bias (Duval & Tweedie, 2000).

2.6 Sensitivity Analysis

The correlation coefficients between BJW and third-party punishment reported in included studies ranged from -0.50 to 0.76, indicating substantial variation. This suggests that meta-analytic results might be influenced by outliers, leading to spurious statistical conclusions (Kepes & Thomas, 2018). To assess outlier impact and result robustness, we conducted two sensitivity analyses using the

leave-one-out method. First, we removed each effect size individually and re-ran the three-level meta-analysis until all effect sizes had been removed once, to gauge outlier effect impact (Dodell-Feder & Tamir, 2018). Second, we removed each original study individually and re-ran the analysis until all studies had been removed once, to gauge outlier study impact (Dodell-Feder & Tamir, 2018).

2.7 Heterogeneity and Moderator Testing

In three-level meta-analysis, we tested three variance sources: sampling error variance (Level 1), within-study effect size variance (Level 2), and between-study effect size variance (Level 3) (Cheung, 2014). We used Q tests to assess overall heterogeneity and one-tailed log likelihood ratio tests for Level 2 and Level 3 variance to determine heterogeneity distribution (Lin et al., 2025; Assink & Wibbelink, 2016). If heterogeneity existed, we followed Higgins et al. (2003) in interpreting I^2 values of 25%, 50%, and 75% as low, medium, and high heterogeneity boundaries, respectively, and conducted moderator analyses to identify heterogeneity sources. For continuous moderators, we used three-level meta-regression to test linear prediction. For categorical moderators, we used dummy coding to test for significant differences between levels (Shen et al., 2025). Moderators included: (1) continuous: female proportion; (2) categorical: age group, sample type, research methodology, BJW type, third-party punishment type, punishment behavior form, cultural background, and violation scenario. Following Card (2012), we required at least 5 effect sizes per categorical moderator level for inclusion in analysis.

2.8 Data Analysis

We conducted meta-analysis using R 4.4.2 with the metafor package (Viechtbauer, 2010), estimating model parameters via restricted maximum likelihood (Viechtbauer, 2010). R code was adapted from programs published by Assink and Wibbelink (2016) and Rodgers and Pustejovsky (2021). Two-tailed $p < 0.05$ was considered significant.

3. Results

3.1 Literature Inclusion and Quality Assessment

This meta-analysis included 46 studies (60 independent samples, 93 effect sizes, 15,772 participants), comprising 10 dissertations and 36 journal articles, with 7 Chinese and 39 English articles spanning 1980–2025. Basic information for included studies is in Appendix A. Quality assessment scores ranged from 6 to 10 (mean = 7.83), indicating good quality (>7 points) (Lin et al., 2025). Overall, included studies were of good quality (see Appendix B).

3.2 Publication Bias and Sensitivity Analysis

The funnel plot (Figure 2 [Figure 2: see original paper]) showed effect sizes roughly evenly distributed in the upper middle and both sides of the plot, visually indicating no severe publication bias. Egger' s test was nonsignificant ($t = 1.71$, $p = 0.087$) with intercept -0.02 (95% CI = $[-0.13, 0.10]$). Overall, no severe publication bias was detected, and trim-and-fill was unnecessary.

Leave-one-out sensitivity analyses showed that removing the effect size reported by He et al. (2021) yielded the lowest correlation ($r = 0.07$, 95% CI = $[0.02, 0.13]$, $p = 0.011$), while removing Russell and Hand (2017) yielded the highest correlation ($r = 0.09$, 95% CI = $[0.04, 0.15]$, $p = 0.002$). Removing each original study individually showed similar patterns, with He et al. (2021) removal producing the lowest correlation and Strömwall et al. (2013) removal producing the highest. In all cases, significance remained unchanged, indicating stable and robust meta-analytic results.

3.3 Main Effect and Heterogeneity Analysis

Three-level meta-analysis revealed a significant positive correlation between BJW and third-party punishment ($r = 0.08$, 95% CI = $[0.03, 0.14]$, $p = 0.006$), representing a small effect size (Gignac & Szodorai, 2016). Variance decomposition showed significant within-study variance (Level 2: $\sigma^2 = 0.01$, $p < 0.001$) and between-study variance (Level 3: $\sigma^2 = 0.03$, $p < 0.001$). Of total variance, sampling variance (Level 1) accounted for 14.60%, within-study variance (Level 2) for 27.35%, and between-study variance (Level 3) for 58.05%. The Q test was significant ($Q(92) = 939.66$, $p < 0.001$, $I^2 = 92.04\%$), indicating substantial heterogeneity and necessitating moderator analysis.

3.4 Moderator Analysis

Moderator analysis revealed several significant effects (Table 1). The self-other dimension of BJW significantly moderated the relationship ($F(1, 55) = 6.20$, $p = 0.013$). Specifically, BJW-self showed a significant positive correlation with third-party punishment ($r = 0.22$, 95% CI = $[0.09, 0.34]$, $p = 0.001$), whereas BJW-other showed no significant correlation ($r = 0.05$, 95% CI = $[-0.03, 0.14]$, $p = 0.236$). Punishment behavior form also significantly moderated the relationship ($F(1, 35) = 10.19$, $p = 0.001$): BJW was positively correlated with monetary punishment ($r = 0.30$, 95% CI = $[0.12, 0.46]$, $p = 0.001$) but not with social punishment ($r = -0.03$, 95% CI = $[-0.13, 0.06]$, $p = 0.500$). Cultural background significantly moderated the relationship ($F(1, 84) = 4.87$, $p = 0.030$): BJW was positively correlated with third-party punishment in collectivist cultures ($r = 0.19$, 95% CI = $[0.09, 0.29]$, $p < 0.001$) but not in individualistic cultures ($r = 0.05$, 95% CI = $[-0.01, 0.12]$, $p = 0.105$). Violation scenario significantly moderated the relationship ($F(2, 45) = 42.99$, $p < 0.001$): BJW was negatively correlated with third-party punishment in sexual violence scenarios ($r = -0.10$, 95% CI = $[-0.15, -0.04]$, $p = 0.002$)

but positively correlated in robbery and corruption scenarios. Further analysis showed significant differences between robbery ($r = 0.14$, 95% CI = $[0.05, 0.24]$, $p = 0.004$) and corruption scenarios ($r = 0.30$, 95% CI = $[0.19, 0.40]$, $p < 0.001$; $Z = 2.04$, $p = 0.041$). Sample type significantly moderated the relationship ($F(3, 89) = 8.59$, $p = 0.035$): BJW was positively correlated with third-party punishment among university students ($M_{age} = 22.01$, $SD = 3.72$; $r = 0.19$, 95% CI = $[0.08, 0.28]$, $p < 0.001$) but not among middle/high school students ($M_{age} = 14.58$, $SD = 0.60$; $r = 0.20$, 95% CI = $[-0.05, 0.43]$, $p = 0.115$), non-student samples ($M_{age} = 36.32$, $SD = 8.58$; $r = 0.17$, 95% CI = $[-0.01, 0.35]$, $p = 0.068$), or unspecified samples ($M_{age} = 30.92$, $SD = 8.30$; $r = 0.02$, 95% CI = $[-0.05, 0.09]$, $p = 0.545$). Age group showed a marginally nonsignificant moderating effect ($F(2, 70) = 5.06$, $p = 0.080$), though between-group comparisons revealed no significant differences between adolescents and adults (adolescent vs. early adult: $Z = -0.55$, $p = 0.580$; adolescent vs. middle adult: $Z = 0.99$, $p = 0.321$), but a significant difference between early and middle adulthood groups ($Z = 2.25$, $p = 0.025$). The explicit-implicit BJW dimension ($F(1, 88) = 0.02$, $p = 0.902$), third-party punishment type ($F(1, 91) = 0.11$, $p = 0.738$), female proportion ($\beta = -0.19$, 95% CI = $[-0.66, 0.29]$, $p = 0.442$), and research methodology ($F(1, 91) = 0.12$, $p = 0.725$) did not significantly moderate the relationship.

4. Discussion

4.1 The Effect of Belief in a Just World on Third-Party Punishment

The significant yet weak positive correlation between BJW and third-party punishment supports the rational strategies hypothesis of just world theory, suggesting that high BJW individuals tend to restore justice through actual behavior rather than cognitive reconstruction, challenging views of BJW as an antisocial or defensive belief system. This can be explained by indirect reciprocity theory and the deontic model of justice. From an indirect reciprocity perspective, BJW is a key cognitive resource promoting third-party punishment. By promising reciprocal norms of "good begets good, evil begets evil," BJW leads punishers to believe their good deeds will be rewarded in the future (Bartholomaeus & Strelan, 2019). This confidence in indirect reciprocity reduces perceived punishment costs (Li et al., 2022; Liu et al., 2025), thereby promoting punishment behavior. From a deontic justice perspective, BJW results from internalizing deservingness norms as an implicit personal contract during childhood (Lerner, 1980), representing an implicit justice motive (Dalbert, 2001) that drives third-party punishment when such beliefs or norms are violated, restoring the "deserved" state stipulated by social norms (Guo et al., 2024; Marshall et al., 2021).

In summary, indirect reciprocity theory and the deontic model explain BJW's facilitative effect on third-party punishment from "economic" and "social" perspectives, respectively. Sensitivity and publication bias analyses show that results are robust and not subject to obvious publication bias. However, significant variance at both within-study (Level 2) and between-study (Level 3) levels

indicates heterogeneity in the main effect, necessitating analysis of potential moderators to further clarify the relationship.

4.2 Moderator Analysis of the BJW-Third-Party Punishment Relationship

Moderator analyses revealed that the self-other dimension of BJW significantly moderated the relationship: BJW-self positively correlated with third-party punishment, whereas BJW-other did not. This suggests the BJW-third-party punishment relationship is not unitary, helping clarify competing hypotheses within just world theory. Specifically, BJW-self emphasizes that "the world is just for me," leading people to believe their good deeds will be rewarded (Bartholomaeus & Strelan, 2019) and to uphold reciprocity in interpersonal interactions (Dalbert, 2012; Sutton et al., 2017). Thus, when witnessing injustice, high BJW-self individuals may punish transgressors based on positive expectations of indirect benefits, supporting the rational strategies hypothesis. In contrast, BJW-other emphasizes that "the world is just for others," prompting people to separate their world from victims' worlds to ignore or avoid injustice threats (Hafer & Rubel, 2015; Lerner, 1980), ultimately weakening motivation to restore justice through third-party punishment, supporting the "multiple-world view" hypothesis.

The explicit-implicit BJW dimension did not significantly moderate the relationship, suggesting no differential effects on third-party punishment. Although explicit and implicit BJW operate at different conscious levels, their overall impact on third-party punishment is consistent, whether through moral intuition or rational deliberation. Previous research also suggests common psychological mechanisms. For example, Guo et al. (2022) found through online surveys and IAT that both explicit and implicit BJW promote altruistic behavior through communal orientation. However, note that this meta-analysis included relatively few implicit BJW effect sizes (only 11.83%), primarily from collectivist cultural contexts, potentially affecting moderator test power. More research on implicit BJW and third-party punishment is needed to verify these findings.

Third-party punishment type did not significantly moderate the relationship. Specifically, BJW was positively correlated with both punishment intention and behavior, with no difference in strength. While this supports the rational strategies hypothesis—that BJW facilitates third-party punishment to restore justice in reality—it may also reflect that participants did not clearly perceive punishment costs within limited experimental timeframes (Shen et al., 2025). Future research should examine whether BJW differentially affects punishment intentions and behaviors when participants clearly perceive varying punishment costs (Chen et al., 2022).

Punishment behavior form significantly moderated the relationship: BJW was positively correlated with monetary punishment but not with social punishment. This provides a punishment-form perspective for clarifying competing hypothe-

ses within just world theory. However, this finding' s generalizability requires cautious interpretation. First, the relationship between BJW and social punishment may be context-dependent (Adolfsson & Strömwall, 2017; Bai et al., 2014; Mikani et al., 2023; Russell & Hand, 2017). Second, existing research has predominantly examined BJW and monetary punishment using third-party game paradigms in collectivist cultural contexts (Chen et al., 2022; He et al., 2021; Yang & Chen, 2022), potentially limiting cultural generalizability. Future research should verify the stability of the BJW-third-party punishment relationship across different cultural backgrounds and violation scenarios.

Cultural background significantly moderated the relationship: BJW was positively correlated with third-party punishment in collectivist but not individualistic cultures. This finding helps reflect on individualistic bias in justice research (Wu et al., 2025) and provides a cultural perspective for clarifying competing hypotheses. However, note that with ongoing modernization, global culture and public psychology show rising individualism and declining collectivism (Huang et al., 2018). China has also shown declining rational individualism and rising utilitarian individualism over the past 40 years (Wu et al., 2024). Against this backdrop of shifting from responsibility-oriented collectivism to self-interest-prioritizing utilitarian individualism, these results provide important insights for moral education based on traditional Chinese culture or chivalrous spirit, cultivating citizens' belief in a just world and 发挥其亲社会功能.

Violation scenario significantly moderated the relationship. Specifically, BJW was negatively correlated with third-party punishment in sexual violence scenarios but positively correlated in robbery and corruption scenarios, with stronger positive correlations in corruption scenarios. This indicates scenario severity is an important factor. Generally, violation severity determines the threat to BJW (Dawtry et al., 2020; Lerner & Simmons, 1966). Compared to robbery and corruption, sexual violence—involving both direct physical harm and long-term psychological trauma—is considered among the most severe scenarios, posing a serious threat to BJW and triggering strong cognitive dissonance and defensive motivation (Hafer & Rubel, 2015; Lerner, 1980). Therefore, in high-threat scenarios, to quickly reduce psychological discomfort and restore justice, bystanders may adopt irrational strategies such as cognitive reconstruction, blaming victims' behavior or character to rationalize the status quo (Ashdown et al., 2024; Hafer & Rubel, 2015; Russell & Hand, 2017), thereby weakening punishment motivation. In contrast, in less severe scenarios, particularly non-violent corruption scenarios with limited BJW threat and lower emotional arousal, people have sufficient cognitive resources to weigh self-interest and social norms (Lerner, 2003, 2023), ultimately showing more rational coping strategies such as punishing transgressors (Bai et al., 2014; Lerner, 2003; Mikani & Rafiee, 2023; Mikani et al., 2023). This provides a scenario-based perspective for clarifying competing hypotheses and highlights the need to consider BJW' s "double-edged sword effect" and scenario dependency in anti-corruption campaigns and #MeToo movements.

Gender did not significantly moderate the relationship, suggesting cross-gender convergence. Although moral orientations differ significantly by gender (Gilligan, 1982; Jaffe & Hyde, 2000), BJW and its functions do not differ significantly by gender (Hammond et al., 2011; Kong et al., 2021; O' Connor et al., 1996; Qiu & Tang, 2025). Thus, gender does not affect the BJW-third-party punishment relationship.

Age group did not significantly moderate the relationship, but sample type did. Specifically, BJW was positively correlated with third-party punishment among university students but not among non-students, middle/high school students, or unspecified samples. Between-group comparisons showed no significant differences between adolescents and adults but significant differences between early and middle adulthood groups. This suggests the developmental trend may follow an inverted U-shape (Kong et al., 2021), with BJW' s facilitative effect on third-party punishment possibly emerging only in early adulthood. This provides a developmental perspective for clarifying competing hypotheses. However, most existing research uses university student samples, with limited examination of children, adolescents, and older adults, restricting interpretability. Future research should use more balanced samples to better understand the lifespan pattern of the BJW-third-party punishment relationship.

Research methodology did not significantly moderate the relationship, indicating no significant difference in correlation strength between questionnaire and experimental studies. However, note that the relationship in experimental studies was unstable and may be influenced by cultural background, violation scenario, and BJW type. Future research should further verify the stability of causal relationships between BJW and third-party punishment.

5. Theoretical Contributions, Limitations, and Future Directions

This three-level meta-analysis integrated domestic and international empirical research on belief in a just world and third-party punishment, examining BJW' s effect and potential moderators. Theoretical contributions include: First, results show a significant yet weak positive correlation between BJW and third-party punishment, supporting the prosocial function of BJW. Second, moderator analyses comprehensively examined potential sources of heterogeneity, testing BJW type, third-party punishment type, punishment behavior form, cultural background, violation scenario, gender, age group, sample type, and research methodology as potential moderators. Results confirmed significant moderation by the self-other dimension, punishment behavior form (social vs. monetary punishment), cultural background, violation scenario, and sample type. These findings clarify competing hypotheses and boundary conditions of just world theory from dimensional, cultural, situational, and developmental perspectives, providing systematic variable perspectives and theoretical extension foundations for future research while offering insights for reflecting on individualistic bias in justice research.

This study has several limitations. First, it only tested competing hypotheses of just world theory from a third-party punishment perspective, neglecting relationships between BJW, third-party compensation, and victim blaming. Future research should comprehensively examine bystander responses to victims and transgressors to better clarify competing hypotheses. Second, previous research suggests that third-party punishment costs (Yin et al., 2019), social distance between bystanders and transgressors (Hu & Halabi et al., 2015), bystander social class, and empathy (Zhou, 2019) may also affect the BJW-third-party punishment relationship, but these variables were insufficiently reported for systematic moderator testing. Future research should increase theoretical development and empirical measurement of such variables to more comprehensively reveal BJW' s mechanisms. Third, some moderators showed uneven effect size distributions. For example, implicit BJW effect sizes were few (only 11.83%), and university student samples dominated, limiting representativeness and stability of moderator results. Interpretations should be cautious. Fourth, the conflict between "social" and "economic" perspectives on BJW remains: whether high BJW individuals punish transgressors based on indirect reciprocity self-interest motives or social norm maintenance responsibility requires further investigation. Finally, only some studies examined mediating mechanisms in the BJW-third-party punishment relationship (Yang & Chen, 2021; Huang et al., 2024; Wu & Cohen, 2017), insufficient for meta-analytic mediation analysis. Future research should expand exploration of BJW' s mechanisms and psychological pathways to deepen theoretical construction and inform intervention practice.

In conclusion, this three-level meta-analysis examined the relationship between belief in a just world and third-party punishment, with three main findings: (1) a significant yet weak positive correlation exists between BJW and third-party punishment; (2) this relationship is moderated by sample type, the self-other dimension of BJW, punishment behavior form, cultural background, and violation scenario—specifically, positive correlations emerged among university students, for BJW-self, for monetary punishment, in collectivist cultures, and in robbery and corruption scenarios (stronger in corruption), but negative correlations appeared in sexual violence scenarios; (3) gender, age group, research methodology, explicit-implicit BJW dimension, and third-party punishment type (behavior vs. intention) did not moderate the relationship.

Note: Figure translations are in progress. See original paper for figures.