The Effect of Third-Party Interventions on Prosocial Behavior: A Three-Level Meta-Analysis

Authors: Shen Yinqi, Cai Yi, Wu Jixia, Wu Jixia

Date: 2025-09-01T00:00:00+00:00

Abstract

As an important force for maintaining social norms, third-party intervention has received extensive attention from researchers for its potential to promote prosocial behavior. To systematically examine the magnitude of its prosocial effects and influencing factors, this study employed a three-level meta-analytic method, integrating 130 effect sizes from 40 studies (totaling 10,289 participants). The main effect test revealed that third-party intervention exerts a medium-to-large promoting effect on prosocial behavior. Moderation analysis indicated that the intensity and probability of third-party intervention significantly influence its promoting effect; specifically, greater intervention intensity and higher probability yield stronger promoting effects on prosocial behavior. However, participant age, gender, type of third-party intervention, format, agent, cost, measurement method of prosocial behavior, and control group setting did not demonstrate significant moderating effects. This study systematically verified the positive impact of third-party intervention on prosocial behavior and clarified a series of key moderating factors, providing implications for subsequent theoretical development and empirical research.

Full Text

The Effects of Third-Party Intervention on Prosocial Behavior: A Three-Level Meta-Analysis

SHEN Yinqi, CAI Yi, WU Jixia*
(School of Education, Soochow University, Suzhou, 215123, China)

Abstract: As an important mechanism for maintaining social norms, third-party intervention has attracted widespread attention from researchers for its potential to promote prosocial behavior. To systematically examine the strength

of this prosocial effect and its influencing factors, this study employed a three-level meta-analytic approach, synthesizing 130 effect sizes from 40 empirical studies involving a total of 10,289 participants. Main effect analyses revealed that third-party intervention exerts a moderately large positive effect on prosocial behavior. Moderator analyses indicated that the intensity and probability of third-party intervention significantly influence its prosocial effects, with stronger and more probable interventions yielding stronger effects. However, no significant moderating effects were observed for participant age, gender, intervention type, form, agent, cost, prosocial behavior measurement paradigm, or control group setting. These findings provide robust evidence for the positive impact of third-party intervention on prosocial behavior and clarify key moderating factors, offering valuable insights for future theoretical and empirical research.

Keywords: third-party intervention, prosocial behavior, three-level meta-analysis, moderating effect
Classification Code: B849: C91

"The sage does not accumulate for himself. The more he gives to others, the more he has for himself." This core prosocial value, deeply rooted in traditional Chinese culture and modern thought, has always been central to maintaining and advancing healthy social relationships (Fehr et al., 2002; Wu et al., 2022). Prosocial behavior encompasses various actions that benefit individuals, groups, organizations, or social welfare (such as helping, donating, volunteering, and cooperating), often at a personal cost to the actor (Bradley et al., 2018; Thielmann et al., 2020). Biological evolution and the internalization of prosocial cultural values have shaped human prosociality and intuition-based prosocial behavior to some extent (Zaki & Mitchell, 2013; Shi & Liu, 2019), yet these factors alone cannot fully explain the emergence of large-scale cooperation and altruism in human societies, particularly prosocial interactions among non-kin or strangers. The development of social norms and the evolution of reciprocal mechanisms provide more comprehensive theoretical foundations for costly prosocial behavior (Glowacki & Lew-Levy, 2022; Rand & Nowak, 2013). Third-party intervention facilitates the enforcement and reinforcement of social norms and the realization of indirect reciprocity (Wu et al., 2022; Guo et al., 2024), serving as an important exogenous mechanism for resolving cooperation dilemmas in social interactions and promoting prosocial behavior (Qin & Wang, 2013). However, existing research has yielded inconsistent conclusions regarding the effectiveness of third-party intervention in promoting prosocial behavior (Mulder et al., 2006; Windrich et al., 2024; Chen et al., 2021), suggesting that the effects of third-party intervention may involve complex boundary conditions. Therefore, it is necessary to integrate existing studies through meta-analysis to systematically examine sources of heterogeneity in effect sizes, further clarifying the relationship between third-party intervention and prosocial behavior and its influencing factors, thereby providing a more comprehensive perspective for subsequent theoretical development and empirical exploration.

1.1.1 The Concept of Third-Party Intervention

Third-party intervention refers to actions taken by an uninvolved third party who, upon observing others'behavior that violates, conforms to, or exceeds social norm expectations, actively punishes, rewards, or compensates the parties involved (Guo et al., 2024; Putz et al., 2016). Specifically, third-party punishment (TPP) involves an impartial party punishing norm violators regarding fairness, cooperation, or other social norms (Chen et al., 2021; Chen & Xin, 2014), including monetary punishment, verbal condemnation, and spreading negative gossip (Festré & Garrouste, 2014; Cui et al., 2017; Chen & Ma, 2011). Third-party reward (TPR) refers to an impartial party rewarding those who comply with norms or exceed normative expectations (Fiedler & Haruvy, 2017; Sutter et al., 2009), including monetary rewards, verbal praise, and spreading positive gossip (Charness et al., 2008; Feinberg et al., 2012). Third-party compensation (TPC) involves an impartial party compensating victims of norm violations (Lotz et al., 2011), including monetary compensation, verbal compensation, and behavioral assistance (Nakashima et al., 2017; Guo et al., 2024).

1.1.2 The Effect of Third-Party Intervention on Prosocial Behavior

Most theoretical and empirical research supports the positive effect of third-party intervention on prosocial behavior. Deterrence theory emphasizes the advantage of third-party punishment in inhibiting norm violations (Guo et al., 2024): when a third-party punisher is present, individuals anticipate potential punishment and violation costs before acting, thereby reducing behaviors that deviate from social norms. For example, this leads to decreased free-riding or defection in social dilemmas and increased contributions to and cooperation with the group (Nakashima et al., 2017; Cui et al., 2017; Chen et al., 2021). Correspondingly, third-party rewards can produce a commitment effect (Fiedler & Haruvy, 2017). When actors anticipate that rewards for prosocial behavior can partially offset their behavioral costs, they become more willing to engage in costly prosocial actions (Sefton et al., 2007; Sutter et al., 2009).

Indirect reciprocity theory also supports the idea that the indirect monetary benefits generated by third-party rewards or punishments increase the attractiveness of individuals engaging in prosocial behavior (Wu et al., 2022). Simultaneously, this theory provides a basis for the effectiveness of social forms of third-party rewards and punishments: individuals may be motivated to establish a good reputation by exhibiting more prosocial behavior, hoping to gain positive social evaluations, avoid the spread of negative gossip, and thereby increase their likelihood of being selected as cooperation partners (Liu & Xin, 2011; Yuan et al., 2016), ultimately obtaining potential indirect benefits in future group interactions or cooperation (Barclay et al., 2021; Roberts et al., 2021). Thus, the reputational costs associated with non-monetary forms of third-party intervention also predispose individuals to engage in prosocial behavior.

Social norm focus theory (Cialdini et al., 1991) provides another layer of support

for the prosocial effects of third-party intervention: punishment and reward are themselves externalizations of social norms, and monetary reward-punishment mechanisms can activate individuals' perception of social norms, while verbal evaluations or gossip dissemination similarly increase norm salience (Eriksson et al., 2021; Shank et al., 2019; Chen, 2022). Third parties maintain norms by punishing or rewarding actors and compensating victims, which makes those intervened upon aware of the behavioral standards that the group endorses or rejects, thereby enhancing their norm compliance (Guo et al., 2024). In other words, third-party intervention can further promote prosocial behavior by enhancing the intervened individuals' perception of both injunctive and descriptive norms. Moreover, this norm focus effect can spill over to bystanders of the intervention or persist into new interaction contexts (Guo et al., 2024; Chen et al., 2021). Additionally, repeated occurrences of third-party intervention may lead actors to undergo reinforced learning of social norms, thereby improving their prosocial behavior in future situations (Fiedler & Haruvy, 2017). For instance, third-party reward or punishment feedback in multi-round interactions leads to continuously increasing cooperation rates among participants (Hou et al., 2019; Nakashima et al., 2017). Based on the above analysis, this study proposes the hypothesis that third-party intervention has a significant positive effect on prosocial behavior.

1.2 Moderating Variables in the Relationship Between Third-Party Intervention and Prosocial Behavior

To further clarify the boundary conditions of third-party intervention effects, this study incorporates multiple moderating variables that existing theoretical and empirical research suggests may cause effect heterogeneity, enabling a more comprehensive understanding of the conditions and mechanisms through which third-party intervention influences prosocial behavior.

1.2.1 Age The prosocial effects of third-party intervention may differ across age stages, a difference closely related to individuals' moral development levels and social cognitive abilities. Children in the norm-learning stage are particularly susceptible to third-party intervention. For example, 6-7-year-olds are in the formation and transformation stage of fairness concepts, and the normative signals conveyed by third-party punishment help deepen their understanding and internalization of social rules, thereby promoting prosocial behavior to a greater extent (Martin et al., 2021; Xiao, 2024). Additionally, although some studies suggest that children under 5 do not yet understand the indirect reputational benefits that gossip may generate (Hill & Pillow, 2006), other evidence indicates that 4-year-olds, when facing third-party gossip threats, already engage in reputation management by exhibiting more prosocial behavior, just like adults (Shinohara et al., 2021). Adolescents are particularly sensitive to social evaluation, with more prominent reputation motives, making them especially likely to be influenced by social forms of third-party intervention (Cui et al., 2017). In summary, age may be a potential variable moderating the relation-

ship between third-party intervention and prosocial behavior.

1.2.2 Gender Women often exhibit higher sensitivity to rewards and punishments (Blackwell, 2000), greater sensitivity to social evaluation (Vanderhasselt et al., 2018), and higher risk aversion (Agnew et al., 2008) than men, and are more responsive to monetary and social feedback (Ding et al., 2017). Therefore, when facing the possibility of third-party punishment, women may be more inclined to engage in prosocial behavior to avoid potential negative consequences (Pablo & Stefania, 2009). Additionally, women tend to focus more on the norm enforcement function of interventions, whereas men are more inclined to weigh gains and losses from a benefit perspective (Burnham, 2018; Qian et al., 2023), and the two genders also differ in their tendencies toward norm enforcement and maintenance (Boschini et al., 2011; Mieth et al., 2017). Overall, gender may influence individuals' responses to third-party intervention, thereby moderating its effect on prosocial behavior, which is why this study includes gender as a moderating variable.

1.2.3 Type of Third-Party Intervention Based on the target of intervention, third-party intervention can be divided into three types: third-party punishment, third-party reward, and third-party compensation (Gummerum et al., 2016; Sutter et al., 2009; Guo et al., 2024). The targets of these three types are respectively the violator, the norm-complier or prosocial actor, and the victim of the violation. This study incorporates these three intervention types as well as pairwise combinations of different types. Regarding the effectiveness of different types of third-party intervention, existing research has not reached consistent conclusions. Early meta-analytic results indicated that punishment and reward have statistically equivalent positive effects on enhancing cooperative behavior (Balliet et al., 2011). However, some studies have proposed different views, suggesting that rewards are more effective than punishments in stimulating individuals' contributions to public goods (Heine & Strobel, 2020; Rand et al., 2009). Other studies have found that third-party rewards are less effective than third-party punishments in inhibiting free-riding behavior (Fiedler & Haruvy, 2017; Zhang, 2019). Similarly, third-party negative gossip has shown stronger effects than positive gossip in promoting cooperative behavior (Wang, 2018). Regarding the effectiveness of third-party compensation, although some researchers argue that third-party compensation is theoretically less deterrent than third-party punishment (Chavez & Bicchieri, 2013), empirical studies have shown that third-party compensation is equally effective as third-party punishment in promoting prosocial behavior (Guo et al., 2024; Wang, 2021). Moreover, researchers and intervention implementers generally believe that combinations of multiple intervention types are superior to single-type interventions (Chen et al., 2015; Hou et al., 2019; Liu, 2018). In summary, the type of third-party intervention may moderate its effect on prosocial behavior.

1.2.4 Form of Third-Party Intervention Third-party intervention forms can be summarized as monetary intervention and social intervention (Asulin et al., 2024; Liu et al., 2010; Chen & Xin, 2014). Monetary intervention typically involves third parties increasing or decreasing other parties' monetary payoffs (Chen et al., 2021), whereas social intervention involves third parties criticizing/praising actors or spreading positive/negative gossip about them (Festré & Garrouste, 2014; Cui et al., 2017). Some research suggests that compared to monetary intervention, which may enhance external motivation while weakening internal motivation, social intervention can place individuals in a moral context, activating their internal and external attributions for cooperative behavior, thereby producing stronger promotion and maintenance of prosocial behavior (Asulin et al., 2024; Liu et al., 2010). However, other studies have shown that praise and criticism are less effective than monetary rewards and punishments in promoting public goods contributions (Zhu, 2009). Thus, different forms of third-party intervention may also lead to differential effects on prosocial behavior.

1.2.5 Agent of Third-Party Intervention Third-party intervention can be implemented by humans or by computer systems. Computer-executed third-party intervention typically operates under pre-programmed rules, with computers providing corresponding intervention feedback based on actors' actual responses (Xiao & Houser, 2011; Liu et al., 2010), whereas human third parties can freely make intervention decisions based on observed behavior (Nakashima et al., 2017; Sutter et al., 2009), a process that often involves the third party' s subjective intentions. As a spontaneous norm maintenance behavior (Guo et al., 2024), third-party intervention executed by humans based on subjective will may be more effective in enhancing individuals' perception of descriptive norms, thereby reducing violations to a greater extent (Chen et al., 2021; Guo et al., 2024). For example, Zhu (2023) manipulated the agent of third-party punishment and found that punishment decisions made by humans significantly affected individuals'fairness perception and emotional experiences, which in turn guided their subsequent prosocial behavior, whereas the effect of computer-agent third-party punishment was relatively weak. However, in terms of intervention accuracy, computer systems are more likely to make precise interventions and provide feedback on violations, potentially producing stronger deterrent effects. Thus, the agent of third-party intervention may also be an important potential moderating variable in the relationship between third-party intervention and prosocial behavior.

1.2.6 Probability, Intensity, and Cost of Third-Party Intervention Third-party intervention probability refers to the likelihood that involved parties will be punished, rewarded, or compensated by a third party. This variable is typically preset by researchers in experiments and includes two manipulation methods: first, pre-setting the probability that a third party will intervene in specific behavioral consequences (Halevy & Halali, 2015; Xiao & Houser, 2011);

second, manipulating the percentage of individuals in a group who are intervened upon (Chen et al., 2015). Third-party intervention intensity reflects the actual loss or gain caused to the intervened party by the intervention (Charness et al., 2008; Guo et al., 2024). Given that social intervention intensity is difficult to quantify, this study only included monetary forms of third-party intervention intensity in the analysis. Additionally, to control for differences in original payoff structures across different game paradigms and enhance cross-study comparability of intervention intensity, this study operationally defined third-party intervention intensity as the ratio of the monetary gain or loss caused by a single minimum-level intervention to the maximum gain from a single violation.

Deterrence theory posits that punishment probability and intensity jointly determine the deterrent power of punishment (Becker, 1968). Theoretically, the higher the probability and intensity of third-party punishment, the stronger its deterrent effect on violations, a view supported by some empirical studies (Windrich et al., 2024; Chen et al., 2015). However, other studies have found that mild third-party punishment is sufficient to effectively promote prosocial behavior, with effects even superior to high-intensity punishment (Kamei, 2020; Chen et al., 2021; Guo et al., 2024). The relationship between punishment probability and prosocial behavior may not be linear but rather an inverted U-shaped curve, where moderate-probability punishment is more effective in promoting prosocial behavior (Qin & Wang, 2013). The effects of other types of third-party intervention may also be moderated by intervention intensity and probability. For example, low-probability, low-intensity third-party rewards may neither produce deterrent effects nor build reciprocal mechanisms, making it difficult to significantly enhance prosocial behavior (Almenberg et al., 2011; Zhang, 2019). High-intensity third-party compensation may play a stronger role in conveying and maintaining fairness norms than low- or medium-intensity compensation, thereby more strongly stimulating prosocial behavior (Wang, 2021).

The cost of third-party intervention can generally be divided into two scenarios: costly and costless. The former involves third parties paying certain monetary or time costs when implementing punishment, reward, or compensation, while the latter involves no resource loss (Chen & Xin, 2014). In the literature included in this study, manipulation of intervention costs was based on monetary forms, specifically whether third parties needed to pay money or tokens when implementing interventions. Intervention cost largely determines third-party intervention intensity and probability (Fiedler & Haruvy, 2017; Chen & Bo, 2016). Individuals often expect costless interventions to be more likely to occur than costly ones, so costless interventions theoretically possess stronger deterrent effects and norm transmission effects (Guo et al., 2024). However, inconsistent research findings indicate that costly interventions are more likely to be perceived as legitimate and altruistic (Raihani & Bshary, 2015), thus promoting prosocial behavior more effectively than costless interventions (Balliet et al., 2011; Kuwabara & Yu, 2017). In summary, third-party intervention intensity, probability, and cost can all serve as potential moderating variables in the effect of third-party intervention on prosocial behavior.

1.2.7 Prosocial Behavior Measurement Paradigm Prosocial behavior measurement paradigms typically include single-interaction tasks and multi-round interaction tasks. The former only requires participants to make a one-time prosocial behavior decision, while the latter allows individuals to make continuous behavioral choices across multiple rounds and may receive third-party intervention feedback in each round to learn about the consequences of their own or others' behavior. Related research indicates that the repetitiveness of third-party intervention helps strengthen the learning process of social norms, thereby more effectively promoting prosocial behavior (Martin et al., 2021; Xiao, 2024). Whether monetary rewards and punishments, social evaluations, or compensation behaviors toward others, all can serve as feedback mechanisms for norm internalization (Raihani et al., 2012; Guo et al., 2024), and multi-round intervention feedback is more conducive to individuals forming stable prosocial behavior patterns (Zheng et al., 2024). Therefore, compared to single interactions, individuals may exhibit higher levels of prosocial behavior in multi-round interactions that include repeated third-party interventions or feedback mechanisms (Bradley et al., 2018).

1.2.8 Control Group Setting When examining the effect of third-party intervention on prosocial behavior, existing experimental studies typically set up two types of control groups: a no-third-party control group and a third-party observer group without intervention rights. Compared to third-party intervention groups that allow third parties to intervene in behavioral outcomes through punishment, reward, or compensation, control groups involve no additional experimental manipulation, while observer groups involve a third party present to observe the interaction process but unable to intervene in any form in the behavioral outcomes. Theoretically, these two types of control groups provide different comparison baselines. In comparison with the no-third-party group, the intervention effect may include the combined effect of both "observer effect" and "intervention effect" (Xiao, 2024). In comparison with the third-party observer group, the third-party intervention effect mainly reflects influence mechanisms beyond the "observer effect," such as monetary gains and losses (Sutter et al., 2009), higher reputational costs (Fehr & Sutter, 2016), and norm reinforcement effects (Chen et al., 2021). Therefore, the effect size measured with the former may be higher than that with the latter, meaning that different control group settings may moderate the magnitude of the effect of third-party intervention on prosocial behavior.

2 Methods

To ensure systematicity and reproducibility, this study followed the PRISMA 2020 statement (Page et al., 2021) and pre-registered the literature screening and data analysis procedures on the Open Science Framework (OSF) after completing literature retrieval (Registration number: 10.17605/OSF.IO/E9BTG).

2.1 Literature Search

This study conducted a comprehensive search of both Chinese and English literature. Chinese literature was primarily searched in CNKI Journal Full-text Database, China Excellent Master' s and Doctoral Dissertations Database, Wanfang Database, and VIP Database. English literature was primarily searched in PubMed, Web of Science, Elsevier, EBSCO, ProQuest, and Google Scholar.

The search procedure was as follows: (1) Search format: "third-party intervention search terms" AND "prosocial behavior search terms" (specific search terms shown in Table 1); (2) Search fields: title and abstract; (3) Final search date: March 2025. A total of 9,324 documents were retrieved. After initial screening to remove duplicates, 5,987 documents were included in the literature library for management; (4) Literature import: Literature data were imported into Zotero reference management software for organization and screening.

Table 1 Literature Search Terms

Third-Party Intervention	Prosocial Behavior
第三方干预、第三方利他、第三方惩罚、第三方制裁、利他性惩罚、第三方奖励、第三方补偿、金钱惩罚/奖励/补偿、社会惩罚/奖励/补偿、言语惩罚/奖励/补偿、社会排斥、第三方帮助、流言、八卦、声誉传播、名声传播	亲社会性、亲社会行为、利他、合作、助人、捐赠、慈善、分享、公平、志愿、信任、互惠、信任游戏、投资游戏、公共物品游戏、最后通牒游戏、独裁者游戏、囚徒困境、社会困境
third-party intervention, third-party punishment, third-party sanction, altruistic punishment, third-party compensation, third-party reward, monetary punishment/reward/compensation, social punishment/reward/compensation, verbal punishment/reward/compensation, social ostracism, third-party help, gossip, reputation spread, reputation transmission	prosociality, prosocial behavior, altruism, cooperation, help, donation, charity, share, fairness, volunteer, trust, trustworthiness, reciprocity, reciprocal, trust game, investment game, public good game, ultimatum game, dictator game, prisoner' s dilemma, social dilemma

2.2 Literature Screening

The literature screening criteria for this study were as follows: (1) Language: Chinese or English, excluding literature in other languages; (2) Study type: Quantitative empirical studies, excluding reviews, meta-analyses, theoretical models, and qualitative research; (3) Since third-party intervention must be implemented through experimental manipulation, included studies must contain a third-party intervention group (i.e., setting up a third party with no direct interest relationship with the actor who can influence others' behavioral outcomes through punishment, reward, or compensation) and at least one control group (i.e., no-third-party condition or third-party observer condition without intervention rights); (4) The measurement target of prosocial behavior must be potential

norm-compliers/violators affected by third-party intervention, excluding studies measuring "violation victims (second parties)" or "intervention implementers (third parties)" ; (5) Included prosocial behavior indicators include but are not limited to: prosocial behavior in dictator games, public goods games, prisoner' s dilemma games, investment games, and prosocial behavioral intentions in social situations; (6) Studies must report sample size, mean, and standard deviation that allow calculation of Hedges' g, or other statistics convertible to g (such as t, ^{2}, F values). Studies reporting only prosocial behavior under third-party intervention conditions or with incomplete data reporting were excluded; (7) When a single article contained multiple independent samples, each independent sample was coded separately; (8) Duplicate publications were excluded, and when duplicate data existed, only the literature with more complete information was selected. The literature screening process is shown in Figure 1 [Figure 1: see original paper].

Figure 1 PRISMA Flowchart

2.3 Literature Coding

Two independent coders coded the included literature according to the coding manual. Basic literature and sample information included: (1) Publication status: formally published journal article (J), dissertation (D), conference paper (C); (2) Mean age and age group of the sample: child (Ch), adolescent (Te), adult (Ad); (3) Female proportion: coded as a continuous variable representing the percentage of female participants in the sample, ranging from 0-100%; (4) Literature ID: author + year; (5) Experiment number: experiment or study number in the original literature.

Third-party intervention characteristics were further coded: (1) Intervention type: third-party punishment (TP), third-party reward (TR), third-party compensation (TC), third-party punishment + reward (TP+TR), third-party punishment + compensation (TP+TC), third-party reward + compensation (TR+TC); (2) Intervention form: monetary intervention (MI), social intervention (SI), monetary + social intervention (MI+SI); (3) Intervention agent: human (H), computer (S); (4) Intervention probability: coded as a continuous variable representing either the probability that a specific behavioral consequence would be intervened upon or the percentage of individuals in a group who were intervened upon, ranging from 0-100%; (5) Intervention intensity: defined as the ratio of the monetary gain or loss from a single minimum-level intervention to the maximum gain from a single violation, calculated as $q = x/y$ (where x = the change in payoff from the minimum intervention level, y = the maximum payoff from a single violation; see Appendix B for calculation methods across different game paradigms). In this study, intervention intensity ranged from $0 < q \leq 3$; (6) Intervention cost: whether the third party needed to pay a monetary cost when implementing the intervention (YES/NO).

Prosocial behavior characteristics were coded: (1) Measurement paradigm:

single-game (SG) vs. repeated-game (RG); (2) Measurement tool: dictator game (DG), public goods game (PGG), prisoner' s dilemma (PD), investment game (IG), or other tools (OT). Additionally, the control group setting was coded: no-third-party control group (CC) vs. third-party observer group without intervention rights (OC).

Coding followed these principles: (1) Each independent sample was coded once, with multiple effect sizes within a study coded individually; (2) When separate group sample sizes were not provided, following Quarmley et al. (2022), the total sample size was divided by the number of groups to estimate each group's sample size; (3) When studies measured multiple variables, each indicator was coded separately. Coding consistency was high, with ICCs for continuous variables ranging from 0.95 to 1.00 and Kappas for categorical variables ranging from 0.86 to 1.00. Discrepancies were resolved through discussion between coders and consultation with the corresponding author.

2.4 Quality Assessment

Each included study was assessed for quality using the Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies. Each item was rated as Yes, No, Cannot Determine (CD), Not Reported (NR), or Not Applicable (NA), with "Yes" scored as 1 and all others as 0. Literature quality was rated as good (total score > 7), fair (total score 5-7), or poor (total score < 5) (Lin et al., 2025). Two independent coders conducted the quality assessment with Kappa = 0.93. Discrepancies were resolved through item-by-item comparison and discussion, with verification against original literature content.

2.5 Effect Size Calculation

This study used Hedges' g as the effect size. Most studies calculated effect sizes from means, standard deviations, and sample sizes, while a few converted F, ^{2}, t, β, or p values to Hedges' g (Harrer et al., 2021). Critical values for small, medium, and large effect sizes were 0.20, 0.50, and 0.80, respectively (Cohen, 1992).

2.6 Model Selection

This study included multiple effect sizes from the same study in most original literature, which violates the assumption of independence in traditional meta-analysis (Cheung, 2014). Three-level meta-analysis accounts for dependency among effect sizes by explaining three sources of variance: sampling variance (Level 1), within-study variance (Level 2), and between-study variance (Level 3) (Cheung, 2014; Van den Noortgate et al., 2013), thereby resolving the non-independence issue while preserving information integrity and improving statistical power (Cheung, 2019). Therefore, this study adopted a three-level random-effects model for analysis.

2.7 Heterogeneity and Moderator Tests

The Q test was used for overall heterogeneity testing to assess whether significant heterogeneity existed among effect sizes. One-tailed log likelihood ratio tests were used to evaluate the significance of within-study and between-study variance to further determine the distribution of heterogeneity (Assink & Wibbelink, 2016; Cheung, 2014; Gao et al., 2024). If heterogeneity existed, moderator tests were conducted to identify its sources. Moderator variables included: (1) continuous variables: female proportion, third-party intervention intensity, third-party intervention probability; (2) categorical variables: age group, intervention type, intervention form, intervention agent, intervention cost, prosocial behavior measurement paradigm, control group setting. Categorical moderators required at least 5 effect sizes per level (Card, 2016).

2.8 Publication Bias Control and Sensitivity Analysis

This study included both published journal articles and unpublished dissertations and conference papers to control for publication bias. Funnel plots, Egger' s regression, and trim-and-fill methods were used to test for publication bias. A symmetric inverted funnel shape indicates small publication bias (Rothstein et al., 2005). Non-significant Egger' s regression results indicate small publication bias (Rodgers & Pustejovsky, 2021). If funnel plots were asymmetric or Egger' s regression was significant, trim-and-fill methods were used to assess the impact of publication bias. If the effect size did not change significantly after trimming and filling, the meta-analytic results were considered minimally affected by publication bias (Duval & Tweedie, 2000).

Sensitivity analysis indicators such as outliers and influence statistics were used to assess result stability. The outlier analysis statistic was studentized deleted residual (SDR), with $|\text{SDR}| > 1.96$ indicating large deviation from predicted mean effect size and thus an outlier, with outlier proportion not exceeding 1/10 of total effect sizes (Viechtbauer, 2010). The influence statistic was DFBETAS, the standardized change in correlation after excluding an effect size. DFBETAS > 1 indicates significant influence of that effect size on the overall effect (Viechtbauer & Cheung, 2010).

2.9 Data Processing

This study conducted meta-analysis using the metafor package (Viechtbauer, 2010) and esc package (Lüdecke, 2019) in R 4.4.3. R code was adapted following tutorials by Assink and Wibbelink (2016) and Harrer et al. (2021).

3 Results

3.1 Study Characteristics

A total of 40 articles were included in the meta-analysis, spanning 2006-2024, containing 57 independent samples, 130 effect sizes, and 10,289 participants.

The number of effect sizes per study ranged from 1 to 10. Effect size numbers for each moderator variable are detailed in Table 2 . Literature quality assessment scores ranged from 5-11, with included literature rated as good ($n = 27$) or fair ($n = 13$). Overall, the included literature was of good quality. Basic information for included literature is in Appendix A.

3.2 Main Effect and Heterogeneity Tests

Using a three-level meta-analytic model to test the main effect of third-party intervention on prosocial behavior, results showed that third-party intervention produced an effect size of $g = 0.73$ ($p < 0.001$, 95% CI [0.57, 0.88]), indicating a significant effect. The Q test revealed significant heterogeneity ($Q(129) = 995.30$, $p < 0.001$). One-tailed log likelihood ratio tests showed significant within-study variance (Level 2) ($\sigma^2 = 0.19$, $p < 0.001$, $I^2 = 50.14\%$) and between-study variance (Level 3) ($\sigma^2 = 0.15$, $p < 0.001$, $I^2 = 39.82\%$). According to Higgins et al. (2003), there was high heterogeneity within studies ($I^2 > 50\%$) and moderate heterogeneity between studies ($I^2 > 25\%$). Therefore, moderator tests were warranted.

3.3 Moderator Tests

This study examined moderating effects of age group (adult/adolescent/child), female proportion, intervention type (punishment/reward/compensation/punishment+reward), intervention form (monetary/social), intervention agent (human/computer), intervention probability, intervention intensity, intervention cost (costly/costless), prosocial behavior measurement paradigm (single-game/repeated-game), and control group setting (no-intervention/observer). For continuous moderators, three-level meta-regression tested linear relationships with effect sizes. For categorical moderators, dummy coding was used to test for differences between levels. Results showed no significant moderating effects for age group ($F(2,115) = 0.47$, $p = 0.627$), female proportion ($F(1,96) = 0.41$, $p = 0.525$), intervention type ($F(3, 126) = 2.09$, $p = 0.106$), intervention form ($F(1, 127) = 0.86$, $p = 0.355$), intervention agent ($F(1, 128) = 2.04$, $p = 0.156$), intervention cost ($F(1, 127) = 0.44$, $p = 0.510$), measurement paradigm ($F(1, 128) = 0.97$, $p = 0.327$), or control group setting ($F(1, 128) = 1.20$, $p = 0.275$). Intervention intensity showed a significant moderating effect, $F(1, 95) = 4.27$, $p = 0.042$, with stronger interventions producing stronger prosocial effects ($b = 0.28$, 95% CI = [0.01, 0.55], $p = 0.042$). Intervention probability showed a marginally significant effect, $F(1, 51) = 3.97$, $p = 0.052$, with higher probability associated with stronger effects ($b = 0.01$, 95% CI [-0.0001, 0.02], $p = 0.052$). Results are shown in Table 2.

Reference groups for categorical moderators were: adult for age group, punishment for intervention type, monetary for intervention form, human for intervention agent, costly for intervention cost, single-game for measurement paradigm, and no-intervention for control group setting.

Table 2 Moderator Tests for the Relationship Between Third-Party Intervention and Prosocial Behavior

Moderator	Intercept/Hedges' g (95% CI)	b (95% CI)	Level 2 Variance	Level 3 Variance
Intercept	0.69 (0.51, 0.87)***	-	0.19***	0.14***
Age Group				
Child	0.59 (0.06, 1.13)*	-0.10 (-0.66, 0.46)	0.20***	0.13***
Adolescent	0.89 (0.46, 1.33)***	0.20 (-0.27, 0.68)	0.21***	0.17**
Adult	0.92 (0.49, 1.35)***	-0.00 (-0.01, 0.01)	0.18***	0.16***
Intervention Intensity	0.68 (0.46, 0.90)***	0.28 (0.01, 0.55)*	0.19***	0.14***
Intervention Probability	0.05 (-0.75, 0.85)	0.01 (-0.0001, 0.02)	0.18***	0.13***
Intervention Type				
Punishment	0.80 (0.61, 0.99)***	-	0.19***	0.14***
Reward	0.55 (0.27, 0.84)***	-0.36 (-0.68, -0.05)*	0.18***	0.13***
Compensation	0.80 (0.36, 1.25)***	-0.28 (-0.85, 0.29)	0.19***	0.15***
Punishment + Reward	0.83 (0.58, 1.12)***	0.00 (-0.39, 0.40)	0.18***	0.14***
Intervention Form				
Monetary	0.76 (0.58, 0.93)***	-	0.19***	0.14***
Social	0.61 (0.34, 0.89)***	-0.14 (-0.44, 0.16)	0.18***	0.14***
Intervention Agent				
Human	0.67 (0.49, 0.84)***	-	0.19***	0.15***

Moderator	Intercept/Hedges' g (95% CI)	b (95% CI)	Level 2 Variance	Level 3 Variance
Computer	0.92 (0.62, 1.12)***	0.25 (-0.10, 0.61)	0.18***	0.13***
Intervention Cost				
Costly	0.80 (0.60, 1.00)***	-	0.19***	0.14***
Costless	0.70 (0.49, 0.92)***	-0.10 (-0.19, 0.38)	0.18***	0.14***
Measurement Paradigm				
Single-game	0.65 (0.42, 0.87)***	-	0.19***	0.15***
Repeated-game	0.79 (0.59, 0.99)***	0.14 (-0.15, 0.43)	0.18***	0.13***
Control Group Setting				
No-intervention	0.78 (0.60, 0.96)***	-	0.19***	0.14***
Observer	0.64 (0.41, 0.86)***	-0.14 (-0.40, 0.12)	0.18***	0.14***

Note: k, number of effect sizes; CI, confidence interval; b, estimated regression coefficient; Level 2 variance = within-study variance; Level 3 variance = between-study variance. $p < 0.05$, **p* < 0.01,** $p < 0.001$.

3.4 Sensitivity Analysis

This study identified 12 outliers (|SDR| > 1.96, Viechtbauer & Cheung, 2010) with effect size IDs 13, 27, 28, 29, 30, 40, 47, 79, 87, 92, 104, and 119. The outlier proportion did not exceed 1/10 of total effect sizes. The DFBETAS plot (Figure 2 [Figure 2: see original paper]) showed that after sequentially removing each effect size, the standardized change in model correlation did not exceed 1 (Cook & Weisberg, 1982; Viechtbauer & Cheung, 2010). Therefore, outlier effect sizes do not substantially influence the results, indicating robust meta-analytic findings.

Figure 2 DFBETAS Plot of Standardized Coefficient Changes After Sequential Effect Size Removal

3.5 Publication Bias Test

The funnel plot (Figure 3 [Figure 3: see original paper]) showed that effect sizes were generally evenly distributed in the upper middle portion and on both sides of the overall effect, though some effect sizes were located in the lower right portion. Egger' s test was significant, $t = 3.49$, $p < 0.001$, with an intercept of 1.92, 95% CI = [0.84, 3.00], indicating some publication bias. Trim-and-fill analysis showed that before removing outliers, adding 30 effect sizes ($k = 160$, original $k = 130$) yielded a corrected effect of $g = 0.52$, 95% CI = [0.38, 0.66], $t = 7.20$, $p < 0.001$. After removing the 12 outliers identified in sensitivity analysis and adding 22 effect sizes ($k = 140$, original $k = 130$), the corrected effect was $g = 0.55$, 95% CI = [0.46, 0.65], $t = 11.36$, $p < 0.001$. Additionally, Rosenthal' s fail-safe $N = 5{,}766$, which exceeds $5 \times k + 10$ ($k = 130$). Therefore, this study is minimally affected by publication bias.

Figure 3 Funnel Plot of Studies on the Relationship Between Third-Party Intervention and Prosocial Behavior

Note: x-axis = Hedges' g scores, y-axis = standard error.

4 Discussion

This study employed a three-level meta-analysis to quantitatively integrate findings from 40 articles, confirming that third-party intervention has a moderately large positive effect on prosocial behavior. The relationship is moderated by third-party intervention intensity and probability. These conclusions deepen understanding of the relationship between third-party intervention and prosocial behavior and provide theoretical insights and implications for future research.

4.1 The Effect of Third-Party Intervention on Prosocial Behavior

The promoting effect of third-party intervention on prosocial behavior can be explained by deterrence theory, indirect reciprocity theory, and social norm focus theory. From a rational actor perspective, norm violations typically occur when individuals want to obtain more benefits (Becker, 1968). Deterrence theory similarly posits that when individuals realize that the violation costs from third-party punishment may exceed the benefits gained from violating norms, violations will not occur (Akers, 1990). From the indirect reciprocity perspective, although individuals who engage in prosocial behavior cannot benefit directly from the current interaction, they may obtain indirect benefits through third-party rewards (Wu et al., 2022). This theory also emphasizes the importance of reputational benefits beyond monetary gains, suggesting that prosocial behavior helps obtain positive third-party evaluations and build positive reputations, which in turn may help individuals benefit in future cooperation and achieve indirect reciprocity (Roberts et al., 2021). Wang' s (2018) empirical research further validated this view, showing that third-party gossip can trigger higher levels of reputational concern, thereby increasing cooperation levels.

Social norms guide people's words and actions, serving as a prerequisite for maintaining prosocial behavior and a factor explaining why humans exhibit prosocial behavior at the expense of their own interests beyond rationality (Gross & Vostroknutov, 2022). Social norm focus theory emphasizes the role of third-party intervention in activating and strengthening social norms (Cialdini et al., 1991; Chen et al., 2015). Punishing violators, rewarding norm-compliers or prosocial actors, and compensating victims are essentially reaffirmations and maintenance of social norms (Guo et al., 2024). Empirical studies by Guo et al. (2024) and Chen et al. (2021) have also confirmed the mediating role of social norm perception in the relationship between third-party intervention and prosocial behavior, indicating that third-party intervention is an effective way to make social norms the focus of individuals' awareness, significantly inhibiting violations and further improving individuals' prosocial levels.

In summary, deterrence theory primarily explains the promoting effect of third-party punishment on prosocial behavior, indirect reciprocity theory emphasizes the monetary and reputational incentive effects of third-party intervention, and social norm focus theory not only supports the prosocial effects of third-party rewards and punishments but also provides a theoretical basis for the effectiveness of third-party compensation. These three theories offer complementary explanatory mechanisms for why different types and forms of third-party intervention can effectively promote prosocial behavior from three dimensions: violation costs, indirect benefits, and norm reinforcement.

4.2 Moderator Analysis of the Relationship Between Third-Party Intervention and Prosocial Behavior

Moderator tests showed significant effects for third-party intervention intensity and marginally significant effects for intervention probability. Specifically, as intervention intensity and probability increase, the promoting effect of third-party intervention on prosocial behavior becomes stronger. Consistent with deterrence theory, higher intensity and probability mean higher costs and risks for violations, thus reducing violation rates (Becker, 1968; Chen et al., 2015). From the indirect reciprocity perspective, the possibility of third-party intervention generating greater indirect benefits attracts individuals to engage more in prosocial behavior to obtain benefits exceeding the costs (Wu et al., 2022). Simultaneously, high-intensity external reward-punishment mechanisms lead to higher cooperation expectations between interaction partners (Balliet et al., 2011; Liu et al., 2010), thereby increasing their own cooperation levels (Lergetporer et al., 2014; Chen & Xin, 2014). Social norm focus theory can also explain this result: higher intervention intensity and probability lead to higher levels of social norm perception, which in turn strengthens individuals' tendency to engage in prosocial behavior, a view confirmed in previous research (Wang, 2021; Chen et al., 2015).

Regarding intervention type, third-party rewards showed weaker effects on prosocial behavior than punishment, while other intervention types did not

differ significantly from punishment. Overall, third-party punishment, reward, compensation, and punishment+reward combinations all significantly promoted prosocial behavior, but rewards were relatively less effective. This result can be explained by prospect theory (Kahneman & Tversky, 1988), which posits that individuals have unequal psychological responses to equivalent "gains" and "losses," showing higher loss sensitivity or "loss aversion." Therefore, compared to the "loss" from punishment, the psychological deterrence produced by third-party rewards is relatively weaker, which may reduce their promoting effect on prosocial behavior (Hou et al., 2019; Zhang, 2019). Additionally, from an information processing and cultural learning perspective, negative information is generally more persuasive and has greater learning value than positive information (Baumeister et al., 2001; Martinescu et al., 2014). Thus, as a negative social signal, third-party punishment may more easily stimulate individuals' understanding and processing of behavioral consequences and strengthen their learning and compliance with social norms, thereby more effectively guiding prosocial behavior (Wang, 2018).

The moderating effect of age was not significant. Humans begin learning social norms from social interactions in early childhood (Bian et al., 2024), and both monetary and social forms of third-party intervention can effectively promote norm compliance (Martin et al., 2021; Shinohara et al., 2021). Although individuals at this stage are still in the norm-learning phase and may have greater room for prosocial behavior improvement, the promoting effect of third-party intervention on prosocial behavior does not vary significantly across age stages. Instead, third-party intervention shows cross-age consistency in its effectiveness in promoting prosocial behavior.

The moderating effect of gender was not significant, indicating cross-gender consistency in the prosocial effects of third-party intervention. Although individuals of different genders may differ in social evaluation and reward-punishment sensitivity (Blackwell, 2000; Vanderhasselt et al., 2018), third-party intervention effectively promotes prosocial behavior for both men and women. Related studies by Guo et al. (2024) and Chen et al. (2021) also showed no significant gender differences in the effect of third-party intervention on prosocial behavior.

The moderating effect of intervention form was not significant, meaning that the effect of third-party intervention on prosocial behavior did not differ by intervention form. Although monetary and social interventions activate different motivations, both material gain and social acceptance are fundamental human motives (Chen & Xin, 2014), so the presence of third-party intervention, regardless of form, can motivate individuals to engage in prosocial behavior. Some previous studies suggested that monetary intervention might crowd out internal motivation and lead to significantly reduced prosocial behavior after intervention removal (Liu et al., 2010; Zhu, 2009). However, this study focused on prosocial behavior during intervention, and results are consistent with these studies' findings during intervention periods, showing that monetary and social interventions are equally effective in promoting prosocial behavior. Addition-

ally, because the number of effect sizes did not meet minimum requirements, the combination of social and monetary intervention could not be included in the meta-analysis (Card, 2016). Therefore, whether simultaneously using multiple forms of third-party intervention produces different effects remains to be explored.

The moderating effect of intervention agent was not significant, meaning that human vs. computer implementation did not produce significantly different effects on prosocial behavior. This result may stem from two reasons: First, participants in some studies may not have psychologically distinguished the agent' s identity. Although computer-executed third-party intervention lacks human subjective intention, it is essentially a feedback mechanism driven by human-preset rules (Liu et al., 2010; Cui et al., 2017). Therefore, participants may view computer agents as "rule enforcers," endowing them with a normative role similar to human interveners. This role equivalence may limit the differential impact of intervention agents on individual behavior, resulting in non-significant moderating effects. Second, individuals' prosocial behavior may be more influenced by intervention outcomes than intentions. In most included experiments, researchers did not manipulate the reasonableness or accuracy of interventions (Charness et al., 2008; Chen et al., 2021). In such cases, participants may focus more on behavioral consequences themselves, with limited room for inferring interveners' subjective intentions, thereby weakening the moderating role of intervention agent.

The moderating effect of intervention cost was not significant; whether third parties needed to pay costs when intervening did not significantly affect the relationship between third-party intervention and prosocial behavior. A possible reason is that intervention cost coding was based on researchers' perspective, which may differ from participants' actual perceptions (Lin et al., 2025). Within limited experimental time, participants may not engage in complex reasoning about the intervention process, focusing instead on proximal factors like intervention outcomes rather than distal factors like intervention costs. This result suggests that intervention cost does not limit the prosocial effects of third-party intervention. However, it should be noted that in studies with costly interventions, cost magnitudes were largely equivalent. Future research could further explore whether prosocial behavior is differentially affected when participants clearly perceive differences in third-party intervention costs.

The moderating effect of measurement paradigm was not significant, meaning that third-party intervention effects did not differ significantly by measurement paradigm. This is consistent with Halevy and Halali (2015), who found that third-party intervention significantly promotes prosocial behavior in both single-game and repeated-game contexts. The effect size was slightly higher for repeated games than single games but not significantly. A possible reason is that third-party intervention in single games already exerts strong influence on behavior, and short-term repeated interactions fail to further enhance this effect. This result demonstrates the robustness of third-party intervention' s

promoting effect on prosocial behavior and provides insights for future research —longitudinal designs could explore whether third-party intervention produces more significant promoting effects over longer time spans.

The moderating effect of control group setting was not significant. Whether using a no-third-party group or a third-party observer group without intervention rights as the control condition, third-party intervention significantly enhanced individuals' prosocial behavior levels. The difference in effect sizes between the two control group settings was not significant, possibly because "observability" itself is insufficient to constitute a strong reputation activation mechanism. Research has pointed out that whether observers can effectively activate actors' reputation motives depends on the existence of direct or indirect reciprocity possibilities between them (Bradley et al., 2018). In third-party observer groups, the third party, as an uninterested party, has no actual reciprocal relationship with participants (Chen et al., 2021), and the reputation concern triggered by their presence is limited. Therefore, participants' prosocial behavior under this condition may be similar to that under no-third-party conditions, limiting the moderating role of control group setting. Additionally, this result demonstrates the robustness of third-party intervention effects, showing that its promoting effect on prosocial behavior does not vary significantly with different control group settings.

5 Theoretical Contributions, Limitations, and Future Directions

5.1 Theoretical Contributions

First, this study reveals, through three-level meta-analysis, that third-party intervention has a moderately large positive effect on prosocial behavior, further validating its effectiveness as a norm maintenance mechanism in promoting prosocial behavior. Second, this study systematically incorporated and tested multiple key moderating variables including age, gender, intervention type, agent, form, probability, intensity, cost, measurement paradigm, and control group setting, comprehensively exploring potential sources of effect heterogeneity. Moderator analysis not only confirms the overall robustness of third-party intervention effects but also identifies that higher intervention intensity and probability enhance its effects. These findings provide support for deterrence theory, social norm focus theory, and indirect reciprocity theory, while revealing the complementarity and applicable boundaries of different theoretical perspectives in explaining third-party intervention mechanisms, and providing a systematic variable framework and theoretical extension basis for future research. Additionally, this study compared intervention effects under different control group settings (no third party vs. third-party observer) and found that third-party intervention significantly enhances prosocial behavior regardless of baseline, indicating that active intervention has greater norm maintenance function than non-intervention or passive observation. In conclusion, this study not only

deepens understanding of the prosocial functions and moderating mechanisms of third-party intervention but also provides theoretical basis and practical insights for how to enhance public cooperation and social welfare through external normative means in reality.

5.2 Limitations and Future Directions

This study has several limitations to be addressed. First, although we included 10 moderating variables as comprehensively as possible, other factors that may influence the relationship between third-party intervention and prosocial behavior warrant further exploration. For example, factors such as the fairness or reasonableness of third-party intervention (Rand et al., 2009; Wu et al., 2022) and the group relationship or social distance between interveners and targets (Harris et al., 2012; Vollan, 2011) have been rarely reported in existing literature and insufficient to support systematic moderator testing. Future research should increase theoretical construction and empirical measurement of such variables to more comprehensively reveal third-party intervention mechanisms.

Second, some included moderators had uneven effect size distributions. For example, third-party compensation had only 5 effect sizes, and age group samples were dominated by adults with relatively few child and adolescent samples, which may limit the representativeness and stability of moderator results. Therefore, interpretations of corresponding results should be made cautiously.

Third, manipulation dimensions of some moderators could be expanded. For instance, because social intervention intensity is difficult to quantify, this study only included monetary forms in intensity moderator analysis. Future research could attempt to construct conversion methods or equivalence models between monetary and social interventions to enhance comparability across forms and strengthen explanatory power and generalizability (Chen & Xu, 2020). Similarly, for intervention cost, current research only covered monetary costs, while whether non-material costs (such as time investment, risk of retaliation) produce differential effects remains to be explored (Chen et al., 2020). Additionally, regarding intervention probability, most included studies focused on high-probability conditions, with relatively scarce data under medium- and low-probability conditions, which may limit full examination of non-linear patterns (e.g., inverted U-shape). Future research should more systematically manipulate intervention probability levels across the full range from low to high to more comprehensively reveal its effect patterns.

Fourth, data missingness in early literature could not be fully overcome. Although we contacted original authors and reviewed supplementary materials, a few early articles still lacked complete descriptive statistics, preventing inclusion of some effect sizes. This highlights the need for future research to strengthen standardized reporting of key statistical indicators to facilitate subsequent review and meta-analysis work.

Finally, although a small number of studies have examined mediating mecha-

nisms between third-party intervention and prosocial behavior, the quantity is insufficient to support meta-analytic examination. Therefore, future research urgently needs to further explore the psychological pathways and mechanisms of third-party intervention to deepen theoretical construction and promote intervention practice.

Conclusion

This study employed three-level meta-analysis to examine the relationship between third-party intervention and prosocial behavior. Results indicate that third-party intervention has a moderately large positive effect on prosocial behavior. The relationship is moderated by intervention intensity and probability but not by age, gender, intervention type, form, agent, cost, measurement paradigm, or control group setting. Overall, the prosocial effects of third-party intervention demonstrate strong stability.

References

Note: The reference list contains both Chinese and English sources. Asterisks indicate references included in the meta-analysis.

Bian, F., Zhang, W., Li, S., & Mu, Y. (2024). The acquisition and development of social norms in early childhood: A social interaction perspective. *Applied Psychology*, 30(4), 323–335. https://doi.org/10.20058/j.cnki.CJAP.023009

Chen, H. (2022). *The effect of third-party punishment on prosocial behavior: The role of social norm demonstration and moral emotions* (Doctoral dissertation). Zhejiang University. https://doi.org/10.27461/d.cnki.gzjdx.2020.004304

Chen, S., & Bo, X. (2016). The influence of fairness and punishment price on third-party punishment demand. *Psychology and Behavior Research*, 14(3), 372–376.

Chen, S. (2011). *Social norm activation: The psychological mechanism of third-party punishment and its impact on cooperative behavior* (Doctoral dissertation). Zhejiang University.

Chen, S., He, Q., & Ma, J. (2015). The effect of third-party punishment on cooperative behavior: An explanation based on social norm activation. *Acta Psychologica Sinica*, 47(3), 389–405.

Chen, S., Hu, H., & Yang, S. (2020). Payment and retaliation: The influence of cost form on third-party punishment. *Psychological Science*, 43(2), 416–422. https://doi.org/10.16719/j.cnki.1671-6981.20200222

Chen, S., & Ma, J. (2011). Third-party punishment and social norm activation: The role of social responsibility and emotion. *Psychological Science*, 34(3), 670–675. https://doi.org/10.16719/j.cnki.1671-6981.2011.03.021

Chen, S., & Yang, S. (2020). The motivation of altruistic punishment. *Advances in Psychological Science*, 28(11), 1901–1910.

Chen, S., Xing, Y., Weng, Y., & Li, C. (2021). The spillover effect of third-party punishment on cooperation: An explanation based on social norms. *Acta Psychologica Sinica*, 53(7), 758–772.

Chen, S., & Xu, Y. (2020). "Benevolent person" or "wise person" : The effect of third-party punishment on the punisher' s reputation. *Acta Psychologica Sinica*, 52(12), 1436–1448.

Chen, X., Zhao, G., & Ye, H. (2014). Forms and functions of punishment in public goods dilemmas. *Advances in Psychological Science*, 22(1), 160–170.

Cui, L., He, X., Luo, J., Huang, X., Cao, W., & Chen, X. (2017). The influence of moral and relational punishment on cooperative behavior of junior high school students in public goods dilemmas. *Acta Psychologica Sinica*, 49(10), 1322–1333.

Guo, Y., Liu, Y., & Cheng, Y. (2024). "Punish the past to prevent future mistakes"and "lead by example": The effect of third-party intervention behavior. *Advances in Psychological Science*, 32(1), 151–161.

Liao, Y., Hong, K., & Zhang, L. (2015). Third-party punishment mechanism and the maintenance of bilateral cooperation order: Experimental evidence from real estate expropriation compensation. *Systems Engineering: Theory & Practice*, 35(11), 2798–2808.

Lin, R., Yu, Q., Hu, T., Zhang, J., Ye, Y., & Lian, R. (2025). A three-level and structural equation model meta-analysis of the relationship between awe and prosocial behavior. *Acta Psychologica Sinica*, 57(4), 631–651.

Liu, G., & Xin, Z. (2011). Reputation mechanisms in indirect reciprocity: Impression, reputation, labels, and their transmission. *Advances in Psychological Science*, 19(2), 233–239.

Liu, X., Ma, J., & Zhu, Y. (2010). Exploring the influence of punishment systems on cooperation in public goods dilemmas from an attribution perspective. *Applied Psychology*, 16(4), 372–377.

Liu, Y. (2018). *Fairness maintenance behavior in gain and loss situations: Third-party punishment and third-party compensation* (Doctoral dissertation). East China Normal University.

Pan, C. (2015). *The role of different reward methods and reward implementers in establishing cooperation order* (Master's thesis). Zhejiang Sci-Tech University.

Shi, R., & Liu, C. (2019). Intuition-based prosociality: Reflections from the social heuristics hypothesis. *Advances in Psychological Science*, 27(8), 1468–1477.

Wang, Q. (2021). *The role of third-party intervention in maintaining and transmitting social norms: The influence of intervention form and intensity* (Master's

thesis). Hebei Normal University. https://doi.org/10.27110/d.cnki.ghsfu.2022.000047

Wang, X. (2018). *The effect of reputation transmission on cooperative behavior: The role of positive and negative information* (Doctoral dissertation). East China Normal University. https://doi.org/10.27149/d.cnki.ghdsu.2018.000009

Xiao, Z. (2024). *The effect of third-party punishment on children' s fair distribution behavior* (Master's thesis). Ludong University. https://doi.org/10.27216/d.cnki.gysfc.2024.000066

Yang, C. (2013). *An experimental study on fairness in urban housing demolition compensation: From the perspective of different demolition methods* (Master' s thesis). China University of Mining and Technology.

Yang, Y., & Hao, J. (2014). The influence of others' altruistic punishment on adults' unfair behavior. (Abstract). In *Proceedings of the 17th National Psychology Academic Conference* (pp. 1216–1218). Beijing.

Yuan, M., Zhang, M., & Kou, Y. (2016). Prosocial reputation and prosocial behavior. *Advances in Psychological Science*, 24(10), 1655–1662.

Zhou, Y., Tu, Q., & Hu, B. (2014). Punishment, social capital, and conditional cooperation: A comparative study based on laboratory and framed field experiments. *Economic Research Journal*, 49(10), 125–138.

Zhang, Y., & Lin, D. (2015). Social preferences, reward-punishment mechanisms, and effective provision of public goods: A study based on an experimental method. *South China Journal of Economics*, 44(12), 26–39. https://doi.org/10.19592/j.cnki.scje.2015.12.003

Zhang, Z. (2019). *The asymmetric role of punishment and reward in centralized sanctioning mechanisms* (Master' s thesis). Zhejiang Sci-Tech University. https://doi.org/10.27786/d.cnki.gzjlg.2019.000133

Zheng, H., Chen, R., & Mai, X. (2024). The cognitive neural mechanisms of third-party punishment behavior. *Advances in Psychological Science*, 32(2), 398–419.

Tian, Y. (2016). *An experimental study on fairness preferences: The influence of dominance, punishment context, and fairness perception on fairness preferences* (Master' s thesis). Harbin Engineering University.

Zhu, S. (2023). *A study on the effect of third-party punishment on second-party victims' subsequent norm compliance behavior* (Master' s thesis). East China Normal University. https://doi.org/10.27149/d.cnki.ghdsu.2023.001508

Zhu, Y. (2009). *A study on the influence of reward-punishment systems on cooperative behavior and trust in public goods dilemmas* (Master' s thesis). Zhejiang University.

Agnew, J. R., Anderson, L. R., Gerlach, J. R., & Szykman, L. R. (2008). Who chooses annuities? An experimental investigation of the role of gen-

der, framing, and defaults. *American Economic Review*, 98(2), 418–422. https://doi.org/10.1257/aer.98.2.418

Akers, R. (1990). Rational choice, deterrence, and social learning theory in criminology: The Path Not Taken. *Journal of Criminal Law and Criminology*, 81(3), 653–676.

Almenberg, J., Dreber, A., Apicella, C. L., & Rand, D. G. (2011). Third party reward and punishment: Group size, efficiency and public goods. In N. M. Palmetti & J. P. Russo (Eds.), *Psychology of Punishment* (pp. 73–91). Nova Science Publishers.

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, 12(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

Asulin, Y., Heller, Y., & Munichor, N. (2024). Comparing the effects of non-monetary incentives and monetary incentives on prosocial behavior. *European Economic Review*, 165, 10740. https://doi.org/10.1016/j.euroecorev.2024.104740

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, 137(4), 594–615. https://doi.org/10.1037/a0023489

Barclay, P., Bliege Bird, R., Roberts, G., & Számadó, S. (2021). Cooperating to show that you care: Costly helping as an honest signal of fitness interdependence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1838), 20200292. https://doi.org/10.1098/rstb.2020.0292

Bašić, Z., Bindra, P. C., Glätzle-Rützler, D., Romano, A., Sutter, M., & Zoller, C. (2021). The roots of cooperation (IZA Discussion Paper No. 14467). Institute of Labor Economics (IZA).

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is Stronger than Good. *Review of General Psychology*, 5(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2), 169–217.

Blackwell, B. S. (2000). Perceived sanction threats, gender, and crime: A test and elaboration of power-control theory. *Criminology*, 38(2), 439–488. https://doi.org/10.1111/j.1745-9125.2000.tb00896.x

Boschini, A., Muren, A., & Persson, M. (2011). Men among men do not take norm enforcement seriously. *The Journal of Socio-Economics*, 40(5), 523–529. https://doi.org/10.1016/j.socec.2011.04.001

Bradley, A., Lawrence, C., & Ferguson, E. (2018). Does observability affect prosociality? *Proceedings of the Royal Society B: Biological Sciences*, 285(1875), 20180116. https://doi.org/10.1098/rspb.2018.0116

Burnham, T. C. (2018). Gender, punishment, and cooperation: Men hurt others to advance their interests. *Socius*, 4, 2378023117742245. https://doi.org/10.1177/2378023117742245

Card, N. A. (2016). *Applied meta-analysis for social science research* (Paperback edition). The Guilford Press.

Charness, G., Cobo-Reyes, R., & Jiménez, N. (2008). An investment game with third-party intervention. *Journal of Economic Behavior & Organization*, 68(1), 18–28. https://doi.org/10.1016/j.jebo.2008.02.006

Chavez, A. K., & Bicchieri, C. (2013). Third-party sanctioning and compensation behavior: Findings from the ultimatum game. *Journal of Economic Psychology*, 39, 268–277. https://doi.org/10.1016/j.joep.2013.09.004

Chen, X., Sasaki, T., Brännström, Å., & Dieckmann, U. (2015). First carrot, then stick: How the adaptive hybridization of incentives promotes cooperation. *Journal of the Royal Society, Interface*, 12(102), 20140935. https://doi.org/10.1098/rsif.2014.0935

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, 19(2), 211–229. https://doi.org/10.1037/a0032968

Cheung, M. W.-L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review*, 29(4), 387–396. https://doi.org/10.1007/s11065-019-09415-6

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 24, pp. 201–234). Academic Press. https://doi.org/10.1016/S0065-2601(08)60330-5

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, 1(3), 98–101. https://doi.org/10.1111/1467-8721.ep10768783

Cook, R. D., & Weisberg, S. (1982). Criticism and influence analysis in regression. *Sociological Methodology*, 13, 313–361. https://doi.org/10.2307/270724

Ding, Y., Wang, E., Zou, Y., Song, Y., Xiao, X., Huang, W., & Li, Y. (2017). Gender differences in reward and punishment for monetary and social feedback in children: An ERP study. *Plos One*, 12(3), e0174100. https://doi.org/10.1371/journal.pone.0174100

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Eriksson, K., Strimling, P., Gelfand, M., Wu, J., Abernathy, J., Akotia, C. S., Aldashev, A., Andersson, P. A., Andrighetto, G., Anum, A., Arikan, G., Aycan, Z., Bagherian, F., Barrera, D., Basnight-Brown, D., Batkeyev, B., Belaus, A.,

Berezina, E., Björnstjerna, M., ···Van Lange, P. A. M. (2021). Perceptions of the appropriate response to norm violation in 57 societies. *Nature Communications*, 12(1), 1481.

Fehr, D., & Sutter, M. (2016). Gossip and the efficiency of interactions. *Games and Economic Behavior*, 113, 448–460. https://doi.org/10.1016/j.geb.2018.10.003

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13(1), 1–25. https://doi.org/10.1007/s12110-002-1012-7

Feinberg, M., Wilier, R., Stellar, J., & Keltner, D. (2012). The virtues of gossip: reputational information sharing as prosocial behavior. *Journal of Personality & Social Psychology*, 102(5), 1015–1030. https://doi.org/10.1037/a0026650

Festré, A., & Garrouste, P. (2014). Somebody may scold you! A dictator experiment. *Journal of Economic Psychology*, 45, 141–153. https://doi.org/10.1016/j.joep.2014.09.005

Fiedler, M., & Haruvy, E. (2017). The effect of third party intervention in the trust game. *Journal of Behavioral and Experimental Economics*, 67, 65–74. https://doi.org/10.1016/j.socec.2016.10.003

Gao, S., Yu, D., Assink, M., Chan, K. L., Zhang, L., & Meng, X. (2024). The association between child maltreatment and pathological narcissism: A three-level meta-analytic review. *Trauma, Violence & Abuse*, 25(1), 275–290. https://doi.org/10.1177/15248380221147559

Glowacki, L., & Lew-Levy, S. (2022). How small-scale societies achieve large-scale cooperation. *Current Opinion in Psychology*, 44, 44–48. https://doi.org/10.1016/j.copsyc.2021.08.026

Gross, J., & Vostroknutov, A. (2022). Why do people follow social norms? *Current Opinion in Psychology*, 44, 1–6. https://doi.org/10.1016/j.copsyc.2021.08.016

Gummerum, M., Van Dillen, L. F., Van Dijk, E., & López-Pérez, B. (2016). Costly third-party interventions: The role of incidental anger and attention focus in punishment of the perpetrator and compensation of the victim. *Journal of Experimental Social Psychology*, 65, 94–104. https://doi.org/10.1016/j.jesp.2016.04.004

Guo, Y., Zhao, X., Liu, Y., & Ma, J. (2024). The effect of third-party intervention on normative behavior of bystanders: An explanation of the social norm cueing effect. *Personality and Individual Differences*, 226, 112694. https://doi.org/10.1016/j.paid.2024.112694

Halevy, N., & Halali, E. (2015). Selfish third parties act as peacemakers by transforming conflicts and promoting cooperation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(22), 6937–6942. https://doi.org/10.1073/pnas.1505067112

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. (2021). *Doing meta-analysis with R: A hands-on guide*. Boca Raton, FL and London: Chapman & Hall/CRC Press. https://doi.org/10.1201/9781003107347

Heim, R., & Huber, J. (2019). Leading-by-example and third-party punishment: Experimental evidence. *Journal of Behavioral and Experimental Finance*, 29, 100436. https://doi.org/10.1016/j.jbef.2019.03.009

Heine, F., & Strobel, M. (2020). Reward and punishment in a team contest. *Plos One*, 15(9), e0236544. https://doi.org/10.1371/journal.pone.0236544

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hill, V., & Pillow, B. H. (2006). Children' s understanding of reputations. *The Journal of Genetic Psychology*, 167(2), 137–157. https://doi.org/10.3200/GNTP.167.2.137-157

Hou, G., Wang, F., Shi, J., Chen, W., & Yu, J. (2019). Which is the ideal sanction for cooperation? An experimental study on different types of third-party sanctions. *PsyCh Journal*, 8(2), 212–231. https://doi.org/10.1002/pchj.259

Kahneman, D., & Tversky, A. (1988). *Prospect theory: An analysis of decision under risk* (p. 214). Cambridge University Press.

Kamei, K. (2020). Group size effect and over-punishment in the case of third party enforcement of social norms. *Journal of Economic Behavior & Organization*, 175, 395–412. https://doi.org/10.1016/j.jebo.2018.04.002

Kuwabara, K., & Yu, S. (2017). Costly punishment increases prosocial punishment by designated punishers: Power and legitimacy in public goods games. *Social Psychology Quarterly*, 80(2), 174–193. https://doi.org/10.1177/0190272517703750

Lergetporer, P., Angerer, S., Glätzle-Rützler, D., & Sutter, M. (2014). Third-party punishment increases cooperation in children through (misaligned) expectations and conditional cooperation. *Proceedings of the National Academy of Sciences of the United States of America*, 111(19), 6916–6921. https://doi.org/10.1073/pnas.1320451111

Lotz, S., Okimoto, T. G., Schlösser, T., & Fetchenhauer, D. (2011). Punitive versus compensatory reactions to injustice: Emotional antecedents to third-party interventions. *Journal of Experimental Social Psychology*, 47(2), 477–480. https://doi.org/10.1016/j.jesp.2010.10.004

Lüdecke, D. (2019). esc: Effect size computation for meta analysis (version 0.5.1). Retrieved July 6, 2022, from https://cran.r-project.org/web/packages/esc/

Martin, J. W., Martin, S., & McAuliffe, K. (2021). Third-party punishment promotes fairness in children. *Developmental Psychology*, 57(6), 927–939. https://doi.org/10.1037/dev0001183

Martinescu, E., Janssen, O., & Nijstad, B. A. (2014). Tell me the gossip: The self-evaluative function of receiving gossip about others. *Personality and Social Psychology Bulletin*, 40(12), 1668–1680. https://doi.org/10.1177/0146167214554916

Mieth, L., Buchner, A., & Bell, R. (2017). Effects of Gender on Costly Punishment. *Journal of Behavioral Decision Making*, 30(4), 899–912. https://doi.org/10.1002/bdm.2012

Mulder, L. B., Van Dijk, E., De Cremer, D., & Wilke, H. A. M. (2006). When sanctions fail to increase cooperation in social dilemmas: considering the presence of an alternative option to defect. *Personality and Social Psychology Bulletin*, 32(10), 1312–1324. https://doi.org/10.1177/0146167206289978

Nakashima, N. A., Halali, E., & Halevy, N. (2017). Third parties promote cooperative norms in repeated interactions. *Journal of Experimental Social Psychology*, 68, 212–223. https://doi.org/10.1016/j.jesp.2016.06.007

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ···Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n71

Pablo, B., & Stefania, O. (2009). Third-party punishment is more effective on women: Experimental evidence (The Papers No. 09/08). Universidad de Granada.

Putz, Á., Palotai, R., Csertő, I., & Bereczkei, T. (2016). Beauty stereotypes in social norm enforcement. *Personality and Individual Differences*, 88, 230–235. https://doi.org/10.1016/j.paid.2015.09.025

Qian, Y., Takimoto, Y., Wang, L., & Yasumura, A. (2023). Exploring cultural and gender differences in moral judgment: A cross-cultural study based on the CNI model. *Current Psychology*, 1–11. https://doi.org/10.1007/s12144-023-04662-6

Qin, X., & Wang, S. (2013). Using an exogenous mechanism to examine efficient probabilistic punishment. *Journal of Economic Psychology*, 39, 1–10. https://doi.org/10.1016/j.joep.2013.07.002

Quarmley, M., Feldman, J., Grossman, H., Clarkson, T., Moyer, A., & Jarcho, J. M. (2022). Testing effects of social rejection on aggressive and prosocial behavior: A meta-analysis. *Aggressive Behavior*, 48(6), 529–545. https://doi.org/10.1002/ab.22026

Raihani, N. J., Thornton, A., & Bshary, R. (2012). Punishment and cooperation in nature. *Trends in Ecology & Evolution*, 27(5), 288–295. https://doi.org/10.1016/j.tree.2011.12.004

Raihani, N. J., & Bshary, R. (2015). The reputation of punishers. *Trends in Ecology & Evolution*, 30(2), 98–103. https://doi.org/10.1016/j.tree.2014.12.003

Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science*, 325(5945), 1272–1275. https://doi.org/10.1126/science.1177418

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, 17(8), 413–425. https://doi.org/10.1016/j.tics.2013.06.003

Roberts, G., Raihani, N., Bshary, R., Manrique, H. M., Farina, A., Samu, F., & Barclay, P. (2021). The benefits of being seen to help others: Indirect reciprocity and reputation-based partner choice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1838), 20200290. https://doi.org/10.1098/rstb.2020.0290

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, 26(2), 141−160. https://doi.org/10.1037/met0000300

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, UK: John Wiley & Sons. https://doi.org/10.1002/0470870168

Sefton, M., Shupp, R., & Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry*, 45(4), 671–690. https://doi.org/10.1111/j.1465-7295.2007.00051.x

Shank, D. B., Kashima, Y., Peters, K., Li, Y., Robins, G., & Kirley, M. (2019). Norm talk and human cooperation: Can we talk ourselves into cooperation? *Journal of Personality and Social Psychology*, 117(1), 99–123. https://doi.org/10.1037/pspi0000163

Shinohara, A., Kanakogi, Y., Okumura, Y., & Kobayashi, T. (2021). Children manage their reputation by caring about gossip. *Social Development*, 31(2), 455–465. https://doi.org/10.1111/sode.12548

Stagnaro, M. N., Arechar, A. A., & Rand, D. G. (2017). From good institutions to generous citizens: Top-down incentives to cooperate promote subsequent prosociality but not norm enforcement. *Cognition*, 167, 212–254. https://doi.org/10.1016/j.cognition.2017.01.017

Sutter, M., Lindner, P., & Platsch, D. (2009). Social norms, third-party observation and third-party reward (Working Paper No. 2009-08). University of Innsbruck.

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, 146(1), 30–90. https://doi.org/10.1037/bul0000217

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior*

Research Methods, 45(2), 576–594.

Vanderhasselt, M.-A., De Raedt, R., Nasso, S., Puttevils, L., & Mueller, S. C. (2018). Don' t judge me: Psychophysiological evidence of gender differences to social evaluative feedback. *Biological Psychology*, 135, 29–35. https://doi.org/10.1016/j.biopsycho.2018.02.017

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. https://doi.org/10.1002/jrsm.11

Vollan, B. (2011). The difference between kinship and friendship: (Field-) experimental evidence on trust and punishment. *Journal of Socio-Economics*, 40(1), 14–25. https://doi.org/10.1016/j.socec.2010.10.003

Windrich, I., Kierspel, S., Neumann, T., Berger, R., & Vogt, B. (2024). Enforcement of fairness norms by punishment: A comparison of gains and losses. *Behavioral Sciences*, 14, 39. https://doi.org/10.3390/bs14010039

Wu, J., Balliet, D., & Van Lange, P. A. M. (2016). Reputation management: Why and how gossip enhances generosity. *Evolution and Human Behavior*, 37(3), 193–201. https://doi.org/10.1016/j.evolhumbehav.2015.11.001

Wu, J., Luan, S., & Raihani, N. (2022). Reward, punishment, and prosocial behavior: Recent developments and implications. *Current Opinion in Psychology*, 44, 117–123. https://doi.org/10.1016/j.copsyc.2021.09.003

Xiao, E., & Houser, D. (2011). Punish in public. *Journal of Public Economics*, 95(7–8), 1006–1017. https://doi.org/10.1016/j.jpubeco.2010.11.021

Zaki, J., & Mitchell, J. P. (2013). Intuitive prosociality. *Current Directions in Psychological Science*, 22(6), 466–470. https://doi.org/10.1177/0963721413492764

Appendix B: Meta-Analysis Coding Manual

Basic Literature Information

Variable	Category/Value	Definition/Range
Publication Status	J = Journal article, D = Dissertation, C = Conference paper	Whether the literature is formally published
Literature ID	Author name + year	For 3+ authors, only first author' s name is coded

Variable	Category/Value	Definition/Range
Mean Age	Arithmetic mean of all participants' ages, typically as reported. If only age range is reported, use the midpoint.	Years
Age Group	Child (Ch), Adolescent (Te), Adult (Ad)	Participant developmental stage, as reported or classified by mean age: Child (<12), Adolescent (13-17), Adult (>18)
Female Proportion	0%-100%	Percentage of female participants in final sample (after excluding invalid cases, if reported)
Experiment Number	Experiment or study number in original literature	

Experimental Design

Variable	Category/Value	Definition
Design Type	Between-subjects, Within-subjects	Whether variable manipulation is between- or within-subjects

Independent Variable Characteristics

Variable	Category/Value	Definition
Third-Party Intervention Type	TP = Punishment, TR = Reward, TC = Compensation, TP+TR, TP+TC, TR+TC	Type(s) of intervention: Punishment (for violations), Reward (for compliance/prosocial behavior), Compensation (for victims), or combinations

Variable	Category/Value	Definition
Intervention Form	MI = Monetary, SI = Social, MI+SI	Form of intervention: Monetary (changing payoffs), Social (verbal, gossip, non-material), or both
Intervention Agent	H = Human, S = Computer/Program	Whether implementer is human or computer
Intervention Probability	0%-100%	Likelihood of intervention: either probability of intervening on specific behaviors, or percentage of group members intervened upon
Intervention Intensity	$q = x/y$	Ratio of minimum intervention' s monetary impact (x) to maximum violation payoff (y). See manual for game-specific calculations. $0 < q \leq 3$.
Intervention Cost	YES = Costly, NO = Costless	Whether third party pays monetary cost to intervene

Dependent Variable Characteristics

Variable	Category/Value	Definition
Prosocial Behavior Measurement Paradigm	SG = Single-game, RG = Repeated-game	Whether behavior measured in single or multiple rounds
Prosocial Behavior Measurement Tool	DG = Dictator Game, PGG = Public Goods Game, PD = Prisoner' s Dilemma, IG = Investment Game, OT = Other	Experimental paradigm used to measure prosocial behavior

Variable	Category/Value	Definition
Control Group Setting	CC = No-third-party control, OC = Third-party observer	Control condition: No third party, or third-party observer without intervention rights

Statistical Analysis Data

Variable	Definition
Original Data Page Number	Page where data appear
Test Statistic	Statistical value from sample data (t, F, ^{2}, r, β, Z, H, U, W, etc.)
Numerator df	Degrees of freedom for numerator variance
Denominator df	Degrees of freedom for denominator variance
Test Statistic Value	Specific value of test statistic
Regression Coefficient SE	Standard error of regression coefficient
p-value	Probability of observing current or more extreme result under null hypothesis
95% CI	95% confidence interval
95% CI Upper/Lower	Upper and lower bounds of confidence interval
Reported Effect Size Type	Type of effect size reported in original paper (d, r, ^{2}, etc.)
Reported Effect Size Value	Standardized effect size reported
Experimental Group Sample Size	Number of participants in experimental group
Experimental Group Prosocial Behavior Proportion	Proportion choosing prosocial behavior (for binary-choice games)
Experimental Group Mean	Mean prosocial behavior (for >2 choice options)
Experimental Group SD/SE	Standard deviation/standard error (for >2 options)
Control Group Sample Size	Number of participants in control group
Control Group Prosocial Behavior Proportion	Proportion choosing prosocial behavior in control group
Control Group Mean/SD/SE	Mean, SD, and SE for control group (for >2 options)

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