中国心理学前沿
2025 年 3 月第 7 卷第 3 期

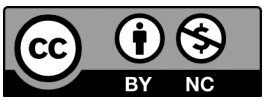
基于社会困境中第三方干预的偏好与其声誉评价：系统综述

张增彤　黄肖玉　赵亚莉　蔡　霞　张　振

河南师范大学，新乡

摘　要 | 本研究明确第三方干预中的选择偏好及其声誉评价，并探究第三方干预及其声誉的关键影响因素。使用混合方法评价工具（MMAT），检索国内外已有研究（2000年—2024年），共3965篇文献（英文3079篇，中文886篇），筛选出59篇，再经过文献质量评价，最终获得满足系统分析条件的研究有53篇。系统综述得出：在第三方干预偏好中，补偿占主导，但惩罚的情境性更加明显；儿童和青少年作为第三方时，其偏好具有发展性特征；成人研究中，偏好差异常受情境驱动；第三方干预偏好是一个受个体特质、情境特征、社会文化及生理心理状态共同作用的复杂决策过程，其核心机制体现在共情正义与成本权衡的互动中。而在声誉评价中，干预类型对声誉评价的影响最值得关注，且遵循“温暖优先”原则，其形成机制可概括为：行为动机——感知维度——文化规范的三阶模型。结论：具有普遍存在补偿偏好现象，其具有较高的声誉。

关键词 | 社会困境；第三方干预；偏好；声誉



1　引言

社会规范的建立与执行是解决社会困境的途径之一，但日常生活中违规行为仍然普遍存在。增加制度约束可以提高人们遵守规范的可能性（Fehr and Schurtenberger，2018），社会干预是一种常见的道德和制度约束行为（Fehr and Fischbacher，2004a），分为第三方补偿和第三方惩罚。除了同样付出自己的一些代价去进行干预的第三方补偿，第三方惩罚在维护社会规范时，对违规者所实施的惩罚行为具有一定的警示作用（Fehr and Fischbacher，2004b；Kanakogi et al.，2022）。第三方干预规范了人类行为，进一步维系并促进了社会公平和社会合作，因而受到研究者们广泛的关注。

与该事件无直接关系的第三方旁观者付出自己的一些代价去惩罚公平的破坏者或者利用自身资源对受害者提供情感支持和物质帮助在学界被称为第三方干预，因其具有较强的利他属性，又被称为第三方利他行为

基金项目：国家社会科学基金项目“教师惩戒和家校沟通影响流动儿童家长对教师信任的机制研究”（24BSH105）。
通讯作者：张振，河南师范大学教育学部心理学院副教授，博士生导师，研究方向：家庭教育研究。
文章引用：张增彤，黄肖玉，赵亚莉，等. 基于社会困境中第三方干预的偏好与其声誉评价：系统综述［J］. 中国心理学前沿，2025，7（3）：408-419.
https://doi.org/10.35534/pc.0703065

（Gummerum et al.，2016；姚佳雯、丁芳，2023）。研究表明，第三方干预者看到公平准则违背事件的不公平程度越大，即接受者受到的对待越不公平，第三方干预者实施的干预强度也会越大（Buckholtz and Marois，2012）。根据指向对象不同，第三方干预行为可以分为第三方惩罚（third-party punishment，TPP）和第三方补偿（third-party compensation，TPC）两种主要形式，即利益无关的第三方愿意自己付出代价，以对造成伤害的违规者做出惩罚（Fehr and Gächter，2002；Fehr and Fischbacher，2004b），或对受到伤害的受害者进行补偿（Leliveld et al.，2012；Lotz et al.，2011）。为了确定人们是倾向于补偿还是惩罚，过去研究直接比较这两种类型的第三方干预：一些研究为惩罚偏好提供了证据（McAuliffe and Dunham，2021；Gummerum et al.，2020；Adams and Mullen，2015；Van Prooijen，2010），而另一些研究认为人们偏好补偿（Arini et al.，2023；Lee and Warneken，2020；Raihani and Bshary，2015）。

而声誉的指向对象是第四方旁观者，即在第三方干预发生之后，个体作为第四方旁观者，通过观察或流言蜚语获得的信息，来了解和预测另一个体（即第三方干预者）未来的行为（Roberts et al.，2021）。先前关于第三方干预的声誉研究主要集中在惩罚上（Fehr and Fischbacher 2004b；Henrich et al.，2006）。过去大多数研究表明，惩罚标志着值得信赖性（McAuliffe and Dunham，2021；Jordan et al.，2016；Raihani and Bshary，2015，Zhang et al.，2023）。然而，尽管从对第三方惩罚的研究中获得了重要的声誉见解，但这绝不是干预声誉的唯一来源。近期研究表明通过补偿对受害者造成的伤害来恢复公平可能会带来额外的有益结果，如增加人际信任（Desmet et al.，2011；Van Doorn et al.，2018）和合作（Fehr and Gächter，2002；Haesevoets et al.，2014），提升社会偏好和特质归因（Lee and Warneken，2020）。

目前人们对第三方干预的偏好与其声誉评价尚不清晰：当面临不公平行为时，人们作为第三方会表现出何种干预偏好？而作为第四方，又会对第三方干预进行怎样的声誉评价？因此，本研究采用系统综述的方法，以明确第三方干预中的选择偏好及其声誉评价，并探究第三方干预及其声誉的关键影响因素。

2 研究方法

本研究采用元分析（PRISMA）指南报告（Page et al.，2021）以及新型综述方式——混合方法研究系统评价（mixed methods systematic review，MMSR）（陈红丽等，2024；廖星 等，2021）进行报告。自变量为第三方干预（惩罚vs补偿vs不干预），因变量为第三方干预偏好和声誉。

2.1 纳入标准

本研究的文献纳入与排除标准如表1所示。

表 1 纳入和排除标准

Table 1 Inclusion and exclusion criteria

纳入标准	排除标准
1. 全文可获取	1. 全文不可获取
2. 无特别时代背景	2. 强调 COVID-19 背景
3. 第三方干预必须包含第三方惩罚和第三方补偿	3. 第三方干预未全部包含第三方惩罚、第三方补偿
4. 研究结果至少包括对第三方干预偏好和声誉的统计分析	4. 研究结果未包括对第三方干预偏好或声誉的统计分析

2.2 文献检索和筛选

在American Psychological Association、Springer Link、Web of Science和Google Scholar 4个数据库中，通过使用第三方干预、第三方反应、惩罚、制裁、补偿、奖励、帮助、不干预（third-party intervention，third-party responses，third-party punishment，third-party sanction，third-party compensation，third-party rewards，third-party helping，do nothing，keep），社会困境、不公平（unfair，injustice，social dilemmas），干预偏好、选择（intervention preference，choice），声誉、感知道德、纯洁性、可信度、慷慨、温暖、能力、正直、仁慈、喜爱、赞许（reputation，perceived morality，purity，trustworthiness，generosity，warmth，competence，integrity，benevolence，like，agreement）等术语来搜索中、英文书面期刊文章的全文。同时，在阅读文后参考文献时利用滚雪球的方法检索文献进行查漏补缺。文献检索的时间范围为2000年到2024年，共检索到文献3965篇，其中英文3079篇，中文886篇。经初筛、审查等阶段后，最终进行文献质量评估的数量为59篇。PRISMA流程图如图1所示。文献筛选由2名研究者按照方案独立进行，有争议的结果由第3位研究者裁决。

https://doi.org/10.35534/pc.0703065

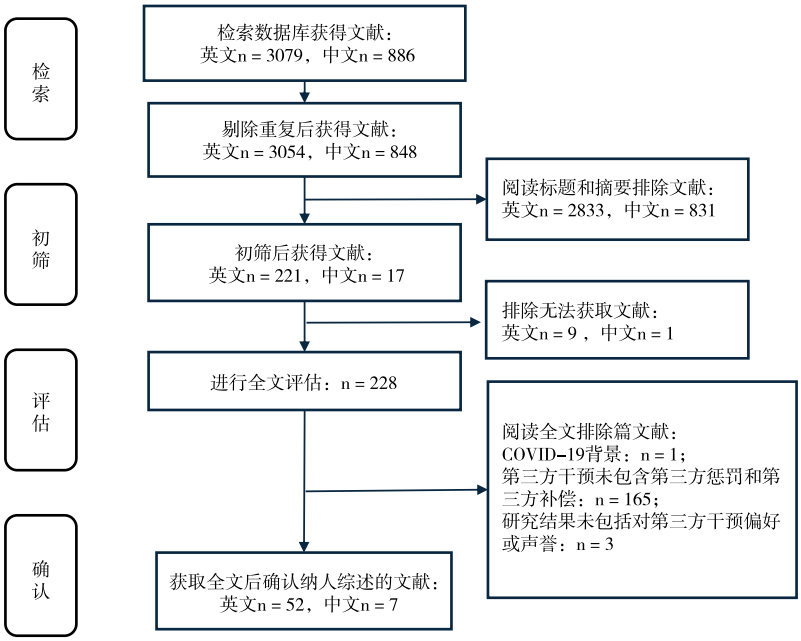


图 1　基于PRISMA指南的文献检索过程

Figure 1　Literature retrieval process based on PRISMA

2.3　文献质量评价

为保证文献质量，本研究根据源文献的特点，使用MMAT进行文献质量评价，由2名评估员独立评估，意见分歧由第3位研究者裁决（廖星 等，2021）。本研究详细记录了每个研究的属性，包括作者（年份）、进行研究的国家、期刊学科、样本量大小、参与者的性别和年龄范围（当研究未提及特定年龄范围且使用平均年龄和标准偏差时，本研究倾向于使用平均年龄和标准偏差）、用于研究干预偏好/声誉的方法和研究设计、干预偏好以及质量评估。此外，本研究记录了每项研究的结论，重点关注第三方干预与干预偏好或声誉之间发现的任何联系的存在和性质，以及最初认为相关的其他发现（例如对结果有重大影响的其他变量），如表2和表3所示。文献评分 ⩾ 60%被认为文献质量较高，本研究共舍去6篇不符合质量评价标准的文章（见表2中文献质量 ⩽ 40%）。

最终获得满足系统分析条件的研究有53篇，其中，中文文献有7篇，英文文献46篇。一共有77个独立样本，其中第三方干预偏好有12070名被试，声誉评价有12929名被试，均包括儿童、青少年、成年人。

表 2　关于第三方干预偏好的质量评价（54篇）

Table 2　Quality evaluation of third-party intervention preferences (Studies of 54)

作者，年份	研究背景			参与者	研究方法		干预偏好	质量评估	
	国家	期刊学科	样本量	性别年龄	方法	设计		MMAT 评分	不符合标准
McAuliffe & Dunham，2021a	M	X	38	F，M（96M ± 12.7M）	Q2	E，A	P	100%	
McAuliffe & Dunham，2021b	M	X	40	F，M（93.5M ± 12.4M）	Q2	E，A	C	100%	
Wang et al.，2022	Z	H	85	F（22.95Y ± 2.19M）	Q2	E，A	P	100%	
Raihani & Bshary，2015	Y，R	O	5241	M，F，D（30.2Y ± 0.2Y）	Q2	E，A	P	100%	
Lee et al.，2021	M	O	120	M（89.64M ± 7.47M）	Q2	E，A	P	100%	
Hu et al ，2020	Y	X	471	M（24.26Y ± 6.02Y）	Q2	E，L	P	100%	
马睿 等，2023	Z	J，X	109	M，F［7Y，10Y］	Q2	E，A	C	100%	
王华根 等，2020	Z	X，S，E	65	M［18Y，29Y］	Q2	E，A	P	100%	

续表

作者，年份	研究背景		参与者		研究方法		干预偏好	质量评估	
	国家	期刊学科	样本量	性别年龄	方法	设计		MMAT 评分	不符合标准
冯佳兴，2020	Z	O	—	D	Q1	R	P，C	60%	1.2，1.4
丁芳 等，2020	Z	J，E	57	F，M（9.22Y ± 0.41Y）	Q2	E，A	C	100%	
Zhen et al.，2020	Z	X	65	F（22.03Y ± 2.10Y）	Q2	E，A	C	60%	3.4，3.5
Jangard et al.，2022	M	H	120	F［18Y，24Y］	Q2	E，A	P，C	80%	3.1
高媛 等，2021	Z	X	117	M，F（21.3Y ± 1.6Y）	Q2	E	P，C	80%	2.4
Wang et al.，2023	Z	X	52	M（23.27Y ± 2.85Y）	Q2	E，A	C	80%	3.4
徐杰 等，2017	Z	X	73	M，F（19.07Y ± 1.16Y）	Q2	E，A	P	80%	4.4
Daniel，2022a	D	X	164	F，M［18Y，33Y］	Q2	E，L	C	100%	
Daniel，2022b	D	X	686	F，M［18Y，82Y］	Q2	E，L	C	100%	
Wiessner，2020	M	O	333	D	Q1	A，R	C	100%	
Kiyonari & Barclay，2008	K	S，X	97	D（大学生）	Q2	E，A	C	80%	2.4
Van Doorn et al.，2018a	M	X，E	308	F，M（31.86Y ± 10.62Y）	Q2	E，A	C	100%	
Van Doorn et al.，2018b	M	X，E	137	F，M（19.44Y ± 2.39Y）	Q2	E，A	C	100%	
Van Doorn et al.，2018c	M	X，E	210	F，M（33.65Y ± 11.01Y）	Q2	E，A	C	100%	
Van Doorn et al.，2018d	M	X，E	202	F，M（31.28Y ± 9.33Y）	Q2	E，A	C	100%	
Van Doorn et al.，2018e	M	X，E	204	F，M（31.74Y ± 9.21Y）	Q2	E，A	N	100%	
Van Doorn et al.，2018f	M	X，E	303	F，M（31.98Y ± 10.71Y）	Q2	E，A	C	100%	
Gary et al.，2006	M	E	144	F，M（大学生）	Q2	E，A	C	80%	2.1
Arini，2023	Y，E，G	X	123	F，M［7Y，11Y］	Q2	E，A	P	60%	2.1，2.2
Lotz et al.，2011	D，M	S，X	178	F，M（22.83Y ± 2.46Y）	Q2	E	C	100%	
Adams & Mullen，2013	M	S，X	694	F，M（31.60Y）	Q2	E，A	C	100%	
Fehr & Fischbacher，2004	H	S，X	73	F，M（20.74Y ± 3.81Y）	Q2	E，A	C，P	100%	
Van Prooijen，2010a	H	S，X	91	F，M（21.22Y ± 3.71Y）	Q2	E，A	P	100%	
Van Prooijen，2010b	H	S，X	106	F，M（21.08Y ± 5.33Y）	Q2	E，A	P，C	100%	
Lee & Warneken，2020a	M	X	82	F，M［60M，118M］	M	E	C	100%	
Lee & Warneken，2020b	M	X	322	F，M［5Y，9Y］	M	A	C	100%	
Van Doorn et al.，2018	H，D	X	402	F，M（20.37Y ± 2.11Y）	Q2	E，A	P	60%	2.1，2.4
Liu et al.，2017	Z	X	169	F，M（21.75Y ± 2.75Y）	Q2	E	C，P	80%	2.1
Yudkin et al.，2016	M，D	X	686	F，M［29Y，39Y］	Q2	E	C，P	80%	2.1
McCall et al.，2014	D	H	33	F，M［45Y，63Y］	Q2	E，A	C，P	60%	2.1，2.4
Gummerum et al.，2016a	Y，H	X	149	F，M［18Y，69Y］	Q2	E，A	P	60%	2.1，2.4
Gummerum et al.，2016b	Y，H	X	140	F，M［18Y，69Y］	Q2	E，A	C	60%	2.1，2.4
Chen et al.，2023	Z	X	483	F，M（大学生）	Q2	E，A	C，P	60%	2.1，2.4

https://doi.org/10.35534/pc.0703065

续表

作者，年份	研究背景		参与者		研究方法		干预偏好	质量评估	
	国家	期刊学科	样本量	性别年龄	方法	设计		MMAT 评分	不符合标准
Sutter et al.，2009	A	X，E	140	D	Q2	E，A	C	60%	2.1，2.2，
Dong et al.，2022	H	X	395	D	Q2	E，A	P	60%	2.1，2.4
Lu & McKeown，2018	Y	X	116	F，M（27.46Y ± 8.72Y）	Q2	E	C，P	80%	2.1
Hechler & Kessler，2022	D	X	302	F，M（39.26Y ± 10.81Y）	Q2	E，A	C，P	80%	2.4
Watanabe，2018	M	X	601	F，M（35.97Y ± 12.05Y）	Q2	E，A	N	60%	2.1，2.4
Côté et al.，2013	M	X	277	F，M（30.48Y ± 10.87Y）	Q2	E，A	P	60%	2.1，2.4
Thulin & Bicchieri，2016	M	S	241	F，M［32Y，34Y］	Q2	E，A	C	40%	2.1，2.2，2.4
Li et al.，2023	Z	O	85	F，M（20.34Y ± 1.70Y）	Q2	E，A	P，C	40%	2.1，2.2，2.4
Giacomantonio & Pierro，2014	E	X	120	F，M（22.26Y ± 3.88Y）	Q2	E，A	P	40%	2.1，2.2，2.4
Rodrigues et al.，2020	Y	O	372	F，M（20.9Y ± 2.01Y）	Q2	E，A	C	20%	2.1，2.2，2.3，2.4
Liu et al.，2019a	M	E	463	F，M（34.2Y ± 10.1Y）	Q2	E，A	C	80%	2.4
Liu et al.，2019b	Z	O	217	F，M（19.07Y ± 1.08Y）	Q2	E，A	P	80%	2.4
Bicchieri & Maras，2022	D	X	56	F，M（26.16Y ± 8.34Y）	Q2	E，A	P	40%	2.1，2.4，2.5
Liu et al.，2018	Z	X	163	F，M（18.79Y ± 0.87Y）	Q2	E，A	P	20%	2.1，2.2，2.4，2.5
Xie et al.，2022	Z	O	60	F，M（18Y，24Y）	Q2	E，A	P，C	80%	2.4
Civai et al.，2019	Y，H	O	38	F，M［19Y，27Y］	Q2	E，A	P，C	60%	2.1，2.4
Koenig & Riley，2017	M	X	89	F，M（23.35 Y ± 7.28Y）	Q2	E，A	P	80%	2.1
Ottone，2005	E	E	48	D	Q2	E，A	C	80%	2.4
Heffner & FeldmanHall，2019	M	X，E	200	F，M（34.9Y ± 10.5Y）	Q2	E，A	C	100%	
Ohtsubo et al.，2018	J	X	87	F，M（19.87Y ± 0.85Y）	Q2	E，A	C	100%	
Will et al.，2013	H	X	183	F，M［9Y，22Y］	Q2	E，A	P，C	80%	3.4
Tang et al.，2024	Z	O	61	F，M（22.2Y ± 1.6Y）	Q2	E，A	C	100%	
Li et al.，2021	Z	X	126	F，M（20.11Y ± 1.61Y）	Q2	E，A	P，C	80%	2.1
Chavez & Bicchieri，2013	M	E，X	196	D（大学生）	Q2	E，A	C	80%	2.1
FeldmanHall et al.，2014	M	X，S，E	540	D	Q2	E，A	P，C	100%	

注：国家—美国 M，中国 Z，英国 Y，德国 D，加拿大 K，意大利 E，哥伦比亚 G，荷兰 H，日本 J，瑞士 R，奥地利 A；期刊学科—教育 J，心理 X，健康 H，社会 S，经济 E，其他 O；性别—M 为女性，F 为男性，D 为不确定性别；年龄 M—是月，Y 是年，年龄范围格式—M（M 岁 / 月 ± SD）、F［xx 岁 / 月，xx 岁 / 月］，分别意为：女性（年龄均值 年 / 月 ± 标准差）、男性［年龄范围 年 / 月］；方法—定性 Q1、定量 Q2、混合 M；设计—综述 R，实验 E，纵向 L，横向 A；干预偏好—惩 P，补 C，不干预 N。（下同）

表 3　关于声誉的质量评价(10篇)

Table 3　Quality evaluation of reputation (Studies of 10)

作者，年份	研究背景			参与者	研究方法		声誉	质量评估	
	国家	期刊学科	样本量	性别年龄	方法	设计		MMAT 评分	不符合标准
Raihani & Bshary，2015	M	O	5241	M，F，D（30.2Y ± 0.2Y）	Q2	E，A	公平感	80%	2.1
吕雪雯，吴航，2024	Z	J	—	M，F［3M，36M］	Q1	R	道德判断	100%	
Fiedler & Haruvy，2017	D	O	449	D	Q2	E，A	可信度	100%	

续表

作者，年份	研究背景			参与者	研究方法		声誉	质量评估	
	国家	期刊学科	样本量	性别年龄	方法	设计		MMAT 评分	不符合标准
Kervyn et al.，2009	M	X	53	F（大学生）	Q2	E，A	温暖、能力	60%	3.4，3.5
Kiyonari & Barclay，2008a	K	S，X	97	D（大学生）	Q2	E，A	社会评价	80%	2.4
Kiyonari & Barclay，2008b	K	S，X	80	D（大学生）	Q2	E，A	社会评价	80%	2.4
Kiyonari & Barclay，2008c	K	S，X	116	D（大学生）	Q2	E，A	社会评价	80%	2.4
Adams & Mullen，2013	M	S，X	694	F，M［31Y，60Y］	Q2	E，A	温暖、能力	100%	
Lee & Warneken，2020a	M	X	80	F，M［5Y，9Y］	Q2	E，A	喜爱程度、社会偏好、特质归因	80%	2.4
Lee & Warneken，2020b	M	X	80	F，M［5Y，9Y］	Q2	E，A	喜爱程度、社会偏好	80%	2.4
Lee & Warneken，2020c	M	X	82	F，M［5Y，9Y］	Q2	E，A	喜爱程度、社会偏好	80%	2.4
Dhaliwal et al.，2024a	M，D，K	X	1002	F（36.18Y ± 12.11Y）；M（34.08Y ± 10.43Y）	Q2	E	道德品质（道德和可信度）	100%	
Dhaliwal et al.，2024b	M，D，K	X	989	F（36.96Y ± 12.33Y）；M（36.21Y ± 12.41Y）	Q2	E	慷慨程度	100%	
Dhaliwal et al.，2024c	M，D，K	X	1634	F（36.41Y ± 11.37Y）；M（34.40Y ± 11.28Y）	Q2	E	赞许性和责备性	100%	
Dhaliwal et al.，2024d	M，D，K	X	1011	F（35.75Y ± 11.71Y）；M（32.47Y ± 10.20Y）	Q2	E	道德品质（道德和可信度）	100%	
Dhaliwal et al.，2024e	M，D，K	X	1014	F（37.17Y ± 11.33Y）；M（35.78Y ± 10.16Y）	Q2	E	道德品质（道德和可信度）	100%	
Heffner & FeldmanHall，2019	M	X，E	200	F，M（34.9Y ± 10.5Y）	Q2	E，A	可信度	100%	
Li et al.，2021	Z	X	107	F，M（19.19Y ± 1.41Y）	Q2	E，A	声誉、友善度、可信度、群体关注度和可敬度维度	80%	2.1

3 研究结果

3.1 第三方干预偏好

本研究发现多数研究中的参与者在面临不公平时，都会做出第三方干预行为，极少数研究发现在惩罚违规者和不干预之间进行迫选时以及干预成本过高时才会选择不干预。第三方干预中的补偿偏好占主导，但惩罚的情境性更加明显。值得注意的是参与者在进行第三方干预时，对“惩罚+补偿”表现出了一定的偏好，尤其在损失情境，表明公平修复需多维策略。

儿童更倾向补偿，但随年龄增长情况趋于复杂化。7~10岁中国儿童在不利不公平情境下偏好补偿，但9岁后惩恶动机逐渐增强（马睿 等，2023）。相比之下，西方儿童（如美国）在第三方角色中更优先惩罚自私行为，例如在独裁者游戏中，儿童更可能对自私分配者采取报复性惩罚而非补偿（Lee and Warneken，2020；Mc Auliffe and Dunham，2021）。此外，在利他惩罚/补偿博弈中，9岁儿童因补偿成本较低更易选择补偿，而14岁及以上青少年更倾向于惩罚排斥者（Will et al., 2013）。总体而言，儿童与青少年更偏好帮助或补偿，其核心动机是减少不平等而非单纯利他（Lee and Warneken，2020）。

成年人对补偿表现出明显偏好，但干预行为也受情境的显著影响。在独裁者游戏中，第三方更倾向于补偿受害者而非惩罚违规者，尤其在补偿能直接修复损失时（Van Doorn et al.，2018）。补偿的确定性（如恢复受害者利益）与惩罚的不确定性（如威慑效果）是驱动这一偏好的关键因素（Van Doorn et al.，2018）。此外，情绪特质（如共情水平）调节干预选择，即高共情者更关注受害者需求（补偿），低共情者则聚焦惩罚违规者（Fehr and Fischbacher，2004）。然而，在严重违规或刑事犯罪中，成年人的惩罚偏好增强。例如，当犯罪者为陌生人且行为意图恶劣时，第三方更倾向于选择惩罚以维护社会规范（Van Prooijen，2010）。这表明成年人的干预策略具有功能分化，即补偿用于恢复受害者损失，惩罚用于强化规范（Koenig and Riley，2017）。

3.2 声誉

补偿行为因符合“利他性社会规范”为普遍被视为“利他信号”，而获得更高声誉。例如，通过公共物品博弈实验发现，在群体合作困境中，实施奖励（如增加他人收益）的第三方因具有“正外部性”（提升整体群体收益）而获得显著更高的声誉评分，奖励者的评分显著高于惩罚者，且被评价为更可信、合作和慷慨（支持间接互惠理论）（Kiyonari and Barclay，2008）。亚当斯和马伦（Adams and Mullen，2013）通过模拟政治竞选

https://doi.org/10.35534/pc.0703065

情境发现，选择补偿受害者（而非惩罚违规者）的候选人因表现出更高的温暖度（如友好、可信赖）而获得更积极的声誉评价，其温暖度评分显著高于惩罚者。值得注意的是，补偿行为的能力维度（如效率、智慧）并未因干预方式不同而产生差异，表明温暖度是声誉评价的核心驱动因素。通过改进的“独裁者游戏”发现，旁观者更倾向奖励帮助者，因其行为被归因为“直接解决问题”和“提升群体价值”（Raihani and Bshary，2015）。达利瓦等人（Dhaliwal et al.，2024）的多项研究进一步表明，补偿者在道德品质（如可信度、慷慨度）上的评分显著优于惩罚者，即使干预成本提高上万倍，这一优势依然稳定。

惩罚行为的声誉效应呈现矛盾性。一方面，其可能传递“维护公平”的信号。例如，惩罚非合作者在公共物品博弈中被合作者视为必要手段，但惩罚者未获得额外声誉优势，甚至被负面评价为“攻击性”或“浪费资源”（Kiyonari and Barclay，2008）。另一方面，发现在强调互惠规范的情境（如信任博弈）中，第三方惩罚者可能因“传递公正”而提升可信度，但其声誉价值受情境背景调节（Heffner and FeldmanHall，2019）。总体而言，惩罚行为需满足“利他归因”条件（如与受害者无利益关联）才能被认可，否则易被怀疑为报复或控制欲（Raihani and Bshary，2015）。

儿童对干预类型的偏好进一步凸显声誉的复杂性。近年来研究发现，5~9岁儿童在被迫选择任务中更偏好帮助者而非惩罚者，且认为帮助者更具“温暖特质”（如同理心）（Lee and Warneken，2020）。8岁以上儿童虽开始理解惩罚的规范性意义，但仍更认可帮助行为，表明积极声誉与亲社会性紧密关联。

3.3 影响因素

3.3.1 影响第三方干预偏好的关键因素

本研究将影响第三方干预偏好的关键因素总结为个体特质、情境特征、社会与文化因素、心理与生理状态4个方面，具体如下。

个体特质会影响干预类型。先前研究指出，高共情者更关注受害者而选择补偿，低共情者则聚焦违规者而偏好惩罚（Fehr and Fischbacher，2004）。同样地，高马基雅维利主义者更倾向惩罚分配者（Liu et al.，2018）。而正义敏感性（Toribio-Flórez，2022）或补偿的确定性（Van Doorn et al.，2018）成为关键调节变量。

情境特征可以改变干预的偏好。高媛等（2021）发现，损失情境下第三方更频繁选择补偿。并且损失情境比收益情境更易激发补偿行为（Liu et al.，2017）。高成本（如需牺牲个人资源）削弱补偿与惩罚行为，但公开情境（如声誉考量）促使第三方更倾向补偿（Li et al.，2021）。另外，声誉的性别差异显著，即女性在高成本下更倾向放弃干预，而男性更可能坚持惩罚以维护“强硬”形象（Jangard et al.，2022）。

第三方干预偏好受到社会与文化因素的影响。徐杰等（2017）发现，第三方对陌生人的惩罚强度显著高于朋友，但对补偿行为无显著差异，表明社会距离通过公平感知间接影响干预选择。集体主义文化可能强化补偿偏好。例如，中国传统“重义扶弱”价值观促使儿童优先补偿受害者（马睿 等，2023）。而西方文化中的个人主义倾向可能强化惩罚动机（McAuliffe and Dunham，2021）。此外，群体身份在内/外群体情境中调节干预行为，即第三方对内群体成员的不公平行为更宽容，倾向于补偿受害者；对外群体成员则更严厉，偏好惩罚（Liu et al.，2018）。

心理与生理状态会影响第三方干预偏决策。研究发现，道德愤怒（如对违规者的义愤）正向影响惩罚行为（Lotz et al.，2011）。第三方与受害者的情感亲近度决定干预类型，即情感疏远时惩罚偏好明显，情感亲近时补偿更受青睐（Van Prooijen，2010）。在快速决策情境下，直觉反应加剧群体偏见，导致对外群体成员的更严厉惩罚（Yuukin et al.，2016）。急性压力降低惩罚意愿，同时提升帮助意愿，可能与应激激素（如皮质醇）对前额叶功能的抑制有关（Wang et al.，2023）。酒精使用障碍患者表现出亲社会决策的减少，包括惩罚与补偿行为（Jangard et al.，2022），提示神经奖赏系统的功能损伤可能削弱干预动机。

以往研究中很值得关注的是个人特质中，年龄通过公平感知与认知发展影响干预偏好（Will et al.，2013）。李与沃纳肯（Lee and Warneken，2020）发现，儿童随年龄增长逐渐偏好惩罚，但补偿仍是其首选策略。成人的惩罚偏好则更多受理性驱动，如在严重犯罪中，成人更倾向惩罚以维护司法公正（Van Prooijen，2010）。

3.3.2 影响声誉评价的关键因素

本研究将影响第三方干预偏好的关键因素总结为干预动机、感知维度、社会规范、文化背景4个方面，具体如下。

（1）第三方行为的动机归因显著影响声誉。若补偿行为被归因为利他（如“想帮助弱者”），其声誉增益更强（Raihani and Bshary，2015）；而惩罚行为需满足“无利益关联”条件才被视为利他（吕雪雯、吴航，2024）。反之，若惩罚被归因为报复或权力欲，其声誉可能受损（Dhaliwal et al.，2024）。例如，在公共物品博弈实验中发现，奖励者因增加他人收益而显著提高群体合作水平，其声誉评分高于惩罚者（Kiyonari and Barclay，2008）。奖励符合“利他性社会规范”，被视为促进群体利益的策略。惩罚因被感知为攻击性行为而损害声誉。这再次印证了以往研究中惩罚者的评分显著低于奖励者，甚至低于未干预者（Kiyonari and Barclay，2008）。惩罚可能破坏“群体和谐”，尤其在集体主义文化中负面效应被放大。

（2）声誉评价会受感知维度的影响。温暖（如友好、道德）是声誉评价的核心维度，可通过补偿效应抵消能力不足，即补偿者因高温暖度弥补了能力劣势（Kervyn et al.，2009；Adams and Mullen，2013）。能力（如效率、智慧）单独作用有限，需与温暖结合。以往研究发现，惩罚者可能具备高能力（如有效阻止搭便车），但惩罚者会被他人认为温暖度低，而导致声誉评价低（Kiyonari and Barclay，2008）。由此，补偿者因高温暖度获得更高声誉，在不同社会类型中均稳定存在，表明“温暖优先”是普遍准则。

（3）声誉评价也受到社会规范的影响。奖励或补偿符合“促进合作”的社会规范。费德勒和哈鲁维（Fiedler and Haruvy，2017）指出，第三方监控结合奖励可显著提升可信度。无独有偶，奖励者被视为更可信的合作者（Kiyonari and Barclay，2008）；而在强调利他主义的情境中（如独裁者游戏），惩罚的声誉价值下降（Heffner and FeldmanHall，2019）。

（4）文化背景影响声誉评价。在集体主义文化（如中国、日本）中，维护群体和谐被视为核心价值，因此帮助或补偿行为可能因其直接促进群体利益而更受认可。吕雪雯和吴航（2024）的研究显示，中国婴幼儿早期即表现出对亲社会者的积极评价，可能与文化中强调“互助”的规范相关。相反，个人主义文化（如美国、加拿大）可能更重视个人责任与规范执行，因此对惩罚行为的接受度更高。以往研究发现惩罚在西方样本中被部分视为“维护公平”，尽管其声誉仍低于帮助行为（Kiyonari and Barclay，2008）。此外，高权力距离文化（如东亚国家）中，惩罚行为若由权威角色（如领导者）实施，可能更易被归因为“维护秩序”而非“攻击性”。海夫纳和费尔德曼霍尔（Heffner and FeldmanHall，2019）提到第三方惩罚在信任博弈中传递“公正信号”，这一效应可能在重视层级规范的文化中更显著。

4 讨论

4.1 第三方干预偏好

作为社会规范的维护者，第三方对不公平事件的干预偏好呈现出多样性与情境依赖性。现有研究表明，第三方既可能选择惩罚违规者、补偿受害者或不干预，也可能同时采用两种策略。总体而言，补偿偏好可能更普遍，尤其在修复受害者损失具有直接确定性时（Van Doorn et al.，2018）；但惩罚偏好同样显著，尤其在违规行为意图恶劣或社会规范被明显破坏时（Falk et al.，2003）。

儿童作为第三方时，其偏好具有发展性特征。例如，马睿等（2023）发现，中国7~10岁儿童在第三方干预中更倾向于补偿受害者，这可能受“扶弱”文化价值观的影响；但随年龄增长，惩罚动机逐渐增强，9岁左右成为转折点。类似地，元分析表明，儿童普遍偏好帮助行为而非惩罚，其核心动机是减少不平等而非单纯利他（Lee and Warneken，2020）。另外，有研究指出，儿童在解释惩罚动机时更强调威慑而非报复，暗示其干预行为具有复杂的道德认知基础（Arini，2023）。

成人研究中，偏好差异常受情境驱动。例如，在急性压力下，第三方可能减少惩罚意愿，转而增加补偿行为（Wang et al.，2023）；而在损失情境中，补偿率显著高于收益情境（高媛 等，2021）。此外，文化背景与社会距离的调节作用显著：中国传统文化可能强化补偿偏好（马睿 等，2023），而第三方对陌生人所做的不公平行为更易引发惩罚（徐杰 等，2017）。

值得注意的是，近年研究强调“混合干预策略”的效用，例如“补偿+惩罚”在修复公平时可兼顾受害者需求与规范维护（Dhaliwal and Cushman，2021）。此外，文化差异持续影响偏好选择：集体主义文化（如中国）中补偿行为因契合“扶弱”价值观而更普遍，而个人主义文化（如美国）中惩罚的威慑功能被强化（Jangard et al.，2022）。

4.2 声誉

声誉评价的核心机制已从单一能力维度转向温暖度与利他性的协同作用。补偿者因传递友好、可信赖的信号获得更高声誉，而惩罚者常因被感知为攻击性而损害社会评价。此外，文化规范调节声誉机制，即在东亚社会，惩罚者的低声誉效应因群体和谐需求被放大，而奖励行为在集体主义背景下更易触发间接互惠（Zhang et al.，2023）。

显而易见，补偿行为在声誉评价中显示出独特的优势。情境模拟实验表明，补偿受害者（而非惩罚违规者）的第三方因表现出更高的温暖度（如友好、可信赖）而获得更积极的声誉评价，而能力维度（如效率、智慧）则不受干预方式影响（Adams and Mullen，2013）。这表明，第四方对第三方干预的评价可能更关注其情感属性（温暖度）而非工具属性（能力）。

4.3 影响因素

4.3.1 影响第三方干预偏好的关键因素

共情水平与正义敏感性构成了干预偏好的两极驱动。高共情者通过情感联结优先补偿受害者（Fehr and Fischbacher，2004），而高正义敏感性者通过规范维护倾向惩罚违规者（Toribio-Flórez，2022）。值得注意的是，当违规行为意图模糊时，个体特质的调节作用凸显——此时补偿确定性（Van Doorn et al.，2018）或正义敏感性成为主导变量，提示认知闭合需求可能影响干预选择。

损失框架与收益框架对干预行为产生不对称影响：损失情境通过增强受害者的“脆弱性凸显效应”显著提升补偿频率（高媛 等，2021），而高成本干预则引发性别差异化反应。男性更可能为维护“强硬”形象坚持惩

https://doi.org/10.35534/pc.0703065

罚（Jangard et al.，2022），这种差异可能与社会化性别角色对风险承担的塑造有关。

集体主义文化通过“扶弱”价值观强化补偿偏好（马睿 等，2023），而社会距离的调节作用揭示了规范维护与共情反应的竞争机制，即对陌生人更倾向惩罚（徐杰 等，2017），可能源于远距离社会关系中“抽象规范”优先于“具象共情”的认知模式。

急性压力通过皮质醇对前额叶功能的抑制，降低惩罚意愿而提升帮助行为（Wang et al.，2023），这为情绪调节理论提供了神经内分泌证据。道德义愤能直接激活对违规者的制裁动机（高媛 等，2021），提示不同情绪状态可能通过双系统加工模型影响干预决策。

4.3.2 影响声誉评价的关键因素

奖励与补偿通过“正外部性”（Kiyonari and Barclay，2008）和温暖感知（Adams and Mullen，2013）获得正向评价，而惩罚因威胁群体和谐被污名化。值得注意的是，能力维度仅在与温暖结合时发挥作用（Kervyn et al.，2009），说明声誉评价本质上是道德特质推断优先的过程。

集体主义文化放大惩罚的负面声誉（Kiyonari and Barclay，2008），这与中国传统“以和为贵”的伦理观形成共鸣。亲社会倾向在生命早期即已显现（吕雪雯、吴航，2024），提示文化规范可能通过与进化机制的交互作用塑造声誉评价系统。

综上所述，本研究通过整合近二十四年（2000—2024）的研究，从第三方干预偏好、声誉评价及影响因素三个层面揭示了社会困境中第三方干预的决策机制与后果。本研究通过系统综述发现，第三方干预偏好是一个受个体特质、情境特征、社会文化及生理心理状态共同作用的复杂决策过程，其核心机制体现在共情正义与成本权衡的互动中。未来的研究可进一步探索干预偏好的神经机制，并设计跨文化实验以揭示规范内化的深层差异。而声誉评价遵循“温暖优先”原则，其形成机制可概括为：行为动机——感知维度——文化规范的三阶模型。未来研究可进一步探索文化差异的深层机制，以及各类关键影响因素对混合干预策略（如同时使用奖励与惩罚）的声誉效应。

参考文献

[1] 陈红丽，丁晓彤，王青，等. 混合方法研究系统评价的方法学及应用进展[J]. 中华护理杂志，2024，59（21）：2665–2671.

[2] 丁芳，刘颜莹，陈甜甜. 儿童道德义愤的发展及其对第三方公正行为的影响[J]. 心理科学，2020，43（3）：652–658.

[3] 高媛，敖丽红，刘映杰. 经济收益与损失影响下的第三方利他惩罚行为[J]. 华北理工大学学报（社会科学版），2021，21（4）：20–26，48.

[4] 廖星，胡瑞学，李博，等. 混合方法研究评价工具的介绍——MMAT[J]. 中国全科医学，2021，24（31）：4015–4020.

[5] 吕雪雯，吴航. 国外道德判断视角下婴幼儿社会评价研究述评[J]. 早期儿童发展，2024（2）：54–64 .

[6] 马睿，吴南，田莫千，等. 惩恶还是扶弱：7～10岁儿童第三方惩罚的动机[J]. 科学通报，2023，68（17）：2258–2268.

[7] 徐杰，孙向超，董悦，等. 人情与公正的抉择：社会距离对第三方干预的影响[J]. 心理科学，2017，40（5）：1175–1181.

[8] 姚佳雯，丁芳. 第三方利他行为：“惩罚”与“补偿”[J]. 心理技术与应用，2023，11（2）：117–128.

[9] Adams G S，Mullen E. Increased voting for candidates who compensate victims rather than punish offenders[J]. Social Justice Research，2013（26）：168–192.

[10] Adams G S，Mullen E. Punishing the perpetrator decreases compensation for victims[J]. Social Psychological and Personality Science，2015（6）：31–38.

[11] Arini R L，Mahmood M，Bocarejo Aljure J，et al. Children endorse deterrence motivations for third–party punishment but derive higher enjoyment from compensating victims[J]. Journal of Experimental Child Psychology，2024（230）：105630.

[12] Buckholtz J W，Marois R. The roots of modern justice：cognitive and neural foundations of social norms and their enforcement[J]. Nature neuroscience，2012，15（5）：655–661.

[13] Charness G，Cobo–Reyes R，Jim é nez N. An investment game with third–party intervention[J]. Journal of Economic Behavior and Organization，2008，68（1）：18–28.

[14] Desmet P T M，De Cremer D，Van Dijk E. In money we trust? The use of financial compensations to repair trust in the aftermath of distributive harm[J]. Organizational Behavior and Human Decision Processes，2011（114）：75–86.

[15] Falk A，Fehr E，Fischbacher U. On the nature of fair behavior[J]. Economic inquiry，2003，41（1）：20–26.

[16] Fehr E，Fischbacher U. Social norms and human cooperation[J]. Trends in Cognitive Sciences，2004，8（4）：185–190.

[17] Fehr E, Fischbacher U. Third-party punishment and social norms [J]. Evolution and Human Behavior, 2004 (25): 63-87.
[18] Fehr E, Gächter S. Altruistic punishment in humans [J]. Nature, 2002 (415): 137-140.
[19] Fehr E, Schurtenberger I. Normative foundations of human cooperation [J]. Nature human behaviour, 2018, 2 (7): 458-468.
[20] Fiedler, Haruvy. The effect of third party intervention in the trust game [J]. Journal of Behavioral and Experimental Economics, 2017 (67): 65-74.
[21] Gummerum M, López-Pérez B, Van Dijk E, et al. When punishment is emotion-driven: Children's, adolescents', and adults' costly punishment of unfair allocations [J]. Social Development, 2020, 29 (1): 126-142.
[22] Gummerum M, Van Dillen L F, Van Dijk E, et al. Costly third-party interventions: The role of incidental anger and attention focus in punishment of the perpetrator and compensation of the victim [J]. Journal of Experimental Social Psychology, 2016 (65): 94-104.
[23] Haesevoets T, Van Hiel A, Reinders Folmer C, et al. What money can't buy: The psychology of financial overcompensation [J]. Journal of Economic Psychology, 2014 (42): 83-95.
[24] Henrich J, McElreath R, Barr A, et al. Costly punishment across human societies [J]. Science, 2006 (312): 1767-1770.
[25] Jangard S, Lindström B, Khemiri L, et al. Alcohol use disorder displays trait-related reductions in prosocial decision making [J]. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2022, 7 (9): 925-934.
[26] Jordan J J, Hoffman M, Bloom P, et al. Third-party punishment as a costly signal of trustworthiness [J]. Nature, 2016, 530 (7591): 473-476.
[27] Kanakogi Y, Miyazaki M, Takahashi H, et al. Third-party punishment by preverbal infants [J]. Nature Human Behaviour, 2022, 6 (9): 1234-1242.
[28] Kervyn N, Yzerbyt V Y, Judd C M, et al. A question of compensation: the social life of the fundamental dimensions of social perception [J]. Journal of Personality and Social Psychology, 2009, 96 (4): 828.
[29] Kiyonari T, Barclay P. Cooperation in social dilemmas: free riding may be thwarted by second-order reward rather than by punishment [J]. Journal of personality and social psychology, 2008, 95 (4): 826.
[30] Lee Y E, Warneken F. Children's evaluations of third-party responses to unfairness: Children prefer helping over punishment [J]. Cognition, 2020 (205): 104374.
[31] Leliveld M C, van Dijk E, van Beest I. Punishing and compensating others at your own expense: The role of empathic concern on reactions to distributive injustice [J]. European Journal of Social Psychology, 2012, 42 (2): 135-140.
[32] Lotz S, Okimoto T G, Schlösser T, et al. Punitive versus compensatory reactions to injustice: Emotional antecedents to third-party interventions [J]. Journal of Experimental Social Psychology, 2011, 47 (2): 477-480.
[33] McAuliffe K, Dunham Y. Children favor punishment over restoration [J]. Developmental science, 2021, 24 (5): e13093.
[34] Page M J, McKenzie J E, Bossuyt P M, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews [J]. International journal of surgery, 2021 (88): 105906.
[35] Raihani N J, Bshary R. Third-party punishers are rewarded, but third-party helpers even more so [J]. Evolution, 2015, 69 (4): 993-1003.
[36] Roberts G, Raihani N, Bshary R, et al. The benefits of being seen to help others: Indirect reciprocity and reputation-based partner choice [J]. Philosophical Transactions of the Royal Society B: Biological Sciences, 2021, 376 (1838): 20200290.
[37] Toribio-Flórez D. An examination of third-party punishment and its boundaries: Between the lab and real life [D]. Technische Universität München, 2022.
[38] Van Doorn J, Zeelenberg M, Breugelmans S M. An exploration of third parties' preference for compensation over punishment: Six experimental demonstrations [J]. Theory and Decision, 2018, 85 (3/4): 333-351.
[39] Van Prooijen J W. Retributive versus compensatory justice: Observers' preference for punishing in response to criminal offenses [J]. European Journal of Social Psychology, 2010, 40 (1): 72-85.
[40] Wang H, Wu X, Xu J, et al. Acute stress reshapes third-party punishment and help decisions: Behavioral evidence and neurocomputational mechanisms [J].

https://doi.org/10.35534/pc.0703065

bioRxiv，2023（6）.
[41] Wang H，Zhen Z，Zhu R，et al. Help or punishment：acute stress moderates basal testosterone's association with prosocial behavior [J]. Stress，2022，25（1）：179–188.
[42] Zhang Z，Li M，Liu Q，et al. Group membership and adolescents' third–party punishment：a moderated chain mediation model [J]. Frontiers in Psychology，2023（14）：1251276.
[43] Heffner J，FeldmanHall O. Why we don't always punish：Preferences for non–punitive responses to moral violations [J]. Scientific Reports，2019，9（1）：13219.
[44] Ohtsubo Y，Sasaki S，Nakanishi D，et al. Within–individual associations among third–party intervention strategies：Third–party helpers，but not punishers，reward generosity [J]. Evolutionary Behavioral Sciences，2018，12（2）：113.
[45] Will G J，Crone E A，van den Bos W，et al. Acting on observed social exclusion：Developmental perspectives on punishment of excluders and compensation of victims [J]. Developmental psychology，2013，49（12）：2236.
[46] Chavez A K，Bicchieri C. Third–party sanctioning and compensation behavior：Findings from the ultimatum game [J]. Journal of Economic Psychology，2013（39）：268–277.
[47] Li Z，Hu G，Xu L，et al. Third–Party Punishment or Compensation? It Depends on the Reputational Benefits [J]. Frontiers in Psychology，2021（12）：676064.
[48] FeldmanHall O，Sokol–Hessner P，Van Bavel J J，et al. Fairness violations elicit greater punishment on behalf of another than for oneself [J]. Nature Communications，2014，5（1）：5306.
[49] Raihani N J，Bshary R. Third–party punishers are rewarded，but third–party helpers even more so [J]. Trends in Ecology and Evolution，2015，30（2）：98–103.
[50] Bicchieri C，Maras M. Intentionality matters for third–party punishment but not compensation in trust games [J]. Journal of Economic Behavior and Organization，2022（197）：205–220.
[51] Chen J，Lian Z，Zheng J. Self–serving reward and punishment：evidence from the laboratory [J]. Scientific Reports，2023，13（1）：13997.
[52] Civai C，Huijsmans I，Sanfey A G. Neurocognitive mechanisms of reactions to second–and third–party justice violations [J]. Scientific Reports，2019，9（1）：9271.
[53] Côté S，Piff P K，Willer R. For whom do the ends justify the means? Social class and utilitarian moral judgment [J]. Journal of personality and social psychology，2013，104（3）：490.
[54] Dong M，Van Prooijen J W，van Lange P A. Strategic exploitation by higher–status people incurs harsher third–party punishment [M]. Social Psychology，2022.
[55] Giacomantonio M，Pierro A. Individual differences underlying punishment motivation：The role of need for cognitive closure [J]. Social Psychology，2014，45（6）：449–457.
[56] Gummerum M，Van Dillen L F，Van Dijk E，et al. Costly third–party interventions：The role of incidental anger and attention focus in punishment of the perpetrator and compensation of the victim [J]. Journal of Experimental Social Psychology，2016（65）：94–104.
[57] Koenig B L，Riley C M. To what reference point do people calibrate cost–free，third–party punishment? [J]. Personality and Individual Differences，2017（115）：90–98.
[58] Liu Y，Bian X，Hu Y，et al. Intergroup bias influences third–party punishment and compensation：In–group relationships attenuate altruistic punishment [J]. Social Behavior and Personality，2018，46（8）：1397–1408.
[59] Liu Y，Li L，Zheng L，et al. Punish the perpetrator or compensate the victim? Gain vs. Loss context modulate third–party altruistic behaviors [J]. Frontiers in Psychology，2017（8）：2066.
[60] Liu Y，Wang H，Li L，et al. Judgments in a hurry：Time pressure affects how judges assess unfairly shared losses and unfairly shared gains [J]. Scandinavian Journal of Psychology，2019，60（3）：203–212.
[61] Lu T，McKeown S. The effects of empathy，perceived injustice and group identity on altruistic preferences：Towards compensation or punishment [J]. Journal of Applied Social Psychology，2018，48（12）：683–691.
[62] McCall C，Steinbeis N，Ricard M，et al. Compassion meditators show less anger，less punishment，and more compensation of victims in response to fairness violations [J]. Frontiers in behavioral neuroscience，2014（8）：424.
[63] Rodrigues J，Liesner M，Reutter M，et al. It's

costly punishment, not altruistic: Low midfrontal theta and state anger predict punishment [J]. Psychophysiology, 2020, 57(8): e13557.
[64] Sutter M, Lindner P, Platsch D. Social norms, third-party observation and third-party reward [M]. Working Papers in Economics and Statistics, 2009.
[65] Tang Y, Hu Y, Zhuang J, et al. Uncovering individual variations in bystander intervention of injustice through intrinsic brain connectivity patterns [J]. NeuroImage, 2024(285): 120468.
[66] Thulin E W, Bicchieri C. I'm so angry I could help you: Moral outrage as a driver of victim compensation [J]. Social Philosophy and Policy, 2016, 32(2): 146–160.
[67] Watanabe S. (2018). Feeling bad and doing good: Forgiveness through the lens of uninvolved others [D]. University of Illinois at Urbana-Champaign, 2018.
[68] Xie E, Liu M, Liu J, et al. Neural mechanisms of the mood effects on third-party responses to injustice after unfair experiences [J]. Human Brain Mapping, 2022, 43(12): 3646–3661.
[69] Yudkin D A, Rothmund T, Twardawski M, et al. Reflexive intergroup bias in third-party punishment [J]. Journal of experimental psychology: general, 2016, 145(11): 1448.

Third-party Intervention Preferences and Their Reputation Evaluation in Social Dilemmas: A Systematic Review

Zhang Zengtong Huang Xiaoyu Zhao Yali Cai Xia Zhang Zhen

Faculty of Education, Henan Normal University, Xinxiang

Abstract: This study aims to clarify third-party intervention preferences and their corresponding reputation evaluations, while also exploring the key factors influencing both intervention and reputational consequences. Utilizing the Mixed Methods Appraisal Tool (MMAT), a comprehensive review of relevant literature published between 2000 and 2024 was conducted. A total of 3,965 studies were retrieved (3,079 in English and 886 in Chinese), of which 59 were initially selected. Following quality assessment, 53 studies met the criteria for systematic analysis. The review revealed that, among third-party intervention preferences, compensation are predominant, whereas punishment exhibit greater situational specificity. When children and adolescents serve as third-party interveners, their preferences demonstrate developmental characteristics. In adult populations, preference differences are largely context-dependent. Third-party intervention preference is a complex decision-making process shaped by individual traits, situational factors, sociocultural influences, and physiological-psychological states. Its underlying mechanism lies in the dynamic interaction between empathic concern for justice and cost-benefit trade-offs. Regarding reputational evaluation, the type of intervention plays a pivotal role and generally aligns with the "warmth-over-competence" principle. The mechanism of reputation formation can be conceptualized as a three-stage model: behavior motivation → perceived dimensions → cultural norms. In conclusion, there is a widely observed preference for compensation, which is associated with higher reputational value.

Key words: Social dilemmas; Third-party intervention; Preferences; Reputation

https://doi.org/10.35534/pc.0703065