《心理学报》审稿意见与作者回应

题目：第三方惩罚对合作的溢出效应：基于社会规范的解释

作者：陈思静、邢懿琳、翁异静、黎常

第一轮

专家 1 的审稿意见及修改说明

意见 1：显然文章并非首次提出社会规范的范畴对第三方惩罚具有调节作用。然而，在“引言”部分却鲜有有关社会规范和第三方惩罚研究的必要回顾，因此我们不太清楚文章的三个研究目的（同时分别对应三个实验）在哪些地方超越了以往研究。同时在三个实验的“结果与讨论”以及“总讨论”部分，也没有与相关研究加以对比。因此，我相信无论是审稿人还是读者，都会产生以下疑惑：本研究在以往研究的基础上，有哪些进一步的新发现。除非作者阐述清楚这个问题，否则无法评判三个实验的结果。一些相关文献包括但不限于：（1）Lois & Wessa, 2019, Creating sanctioning norms in the labs: The influence of descriptive norms in third-party punishment, Social Influence；（2）Rupp & Bell, 2010, Extending the denotic model of justice: Moral self-regulation in third-party responses to injustice, Business Ethics Qualrterly；（3）陈思静，何铨 & 马剑虹，2015，第三方惩罚对合作行为的影响：基于社会规范激活的解释，心理学报。

修改说明 1：感谢专家富有建设性的意见，这给了我们很大启发。诚如您所言，本文并非首个从社会规范角度来探讨第三方惩罚的研究，现有文献已从不同角度探索了社会规范与第三方惩罚之间的关系，并得出了很多有意义的结论：如 Bicchieri 等（2018）和 Fehr 和 Williams（2018）的研究结果表明惩罚必须与相应的规范结合才能提高受罚者的合作水平；陈思静等（2015）也将社会规范与第三方惩罚结合了起来，并指出第三方惩罚本身即是一种社会规范激活的过程；此外，专家所补充的文献 Lois 和 Wessa（2019）也描述了当第三方惩罚和描述性规范传达了不一致的信息时后者对前者的调节作用。这些研究都强调了第三方惩罚促进合作中不仅有“经济效应”（降低受罚者的报酬），也存在“规范效应”（提示受罚者存在有关合作的规范），换言之，第三方惩罚对合作的促进并非仅仅因为其改变了违规者的收益结构，还因为其提示了相应的社会规范。但由于在主流研究中第三方惩罚总是改变了违规者的收益，因此我们仍然不清楚惩罚的这两种效应是什么关系：是独立起作用还是相互依赖？就

本文作者所知，目前尚无研究考察过这一问题。我们希望在这方面将有关第三方惩罚的研究推进一步，并回答以下问题：当第三方惩罚不足以影响违规者的收益时，是否还能有效地促进合作？如果答案是肯定的，那就说明规范效应并非需要经济效应为前提，而是第三方惩罚的一种独立作用，这对现有研究的结论是一个有力的补充。就我们所知，目前应该还没有研究把惩罚的经济效应和规范效应通过实验进行区分，而本文的第一个创新点就在于通过排除第三方惩罚的经济效应而为其纯粹的规范效应提供了明确的实证证据：即使第三方惩罚无法降低违规收益，依然能提高违规者的合作水平。这正是本文实验 1 的主旨所在。

其次，更为重要的是，我们从实验 1 第三方惩罚的规范效应这个视角出发进一步发现了惩罚提升合作的两种溢出效应。如果第三方惩罚的规范效应仅仅体现在了“受罚者本人在某个特定场景下的合作规范被激活了从而提高了合作水平”，那么我们又陷入了类似用经济角度去解释第三方惩罚作用的困境：如果必须通过惩罚对每个个体在每个场景下进行规范提示，那么社会运行成本会极高，第三方惩罚也就没有存在的意义了。基于上述逻辑，我们推测第三方惩罚的规范提示作用可能存在溢出效应并通过两个实验检验了这个推测：曾经受罚的个体在没有惩罚机制的新情境下是否依然表现出积极的合作行为（纵向溢出效应，实验 2）？以及，旁观而没有亲身经历惩罚的个体是否会在缺乏惩罚机制的情境中表现出合作行为（横向溢出效应，实验 3）？实验结果验证了我们的推测，即确实存在第三方惩罚的两种溢出效应。和先前文献相比，这为我们理解人类社会的长时间（纵向溢出效应）、大规模（横向溢出效应）的合作提供了新的理论思路。在上述发现的基础上，我们进一步比较了为什么基于规范激活的溢出效应观点比基于经济学观点的威慑观点（Gintis & Fehr, 2012）能更好地解释第三方惩罚对人类社会广泛合作的维系作用。据本文作者所知，本文首次基于社会规范激活的角度提出了第三方惩罚的两种溢出效应，并以此探讨了维持人类社会广泛合作的一种潜在机制。这是本文的第二个创新点。

第三，由于社会规范在心理学文献中通常被分为描述性规范和命令性规范（Cialdini et al., 1991），因此我们在实验 2 和 3 中分别检验了两种规范在中介惩罚与合作关系中的机制，并基于现有文献中的相关理论对两种规范机制的异同进行了讨论，进一步加深了对第三方惩罚通过激活社会规范而影响合作行为的机制的理解，从而为现有文献提供了有益补充。这是本文的第三个创新点。

由于我们在行文上存在一定纰漏，没能在前言、讨论和总讨论中清楚地阐述上述问题，从而无法很好地体现出本文的几个创新点，专家的建议启发了我们对整篇文章在该领域知识结构中意义和作用的反思，使我们受益良多。我们重新撰写了前言、3 个实验的讨论和总讨

论，从而使表述更为清楚。由于修改部分较多，不便在这里赘述，烦请专家参看正文标蓝部分，并再次感谢您的宝贵意见！

意见 2：对应于中文摘要也存在以上提及的问题：关于文章从事了什么研究、为什么进行该研究、以及如何进行的该研究的目的性描述不清楚。摘要尽量以“通俗易懂”的方式叙述，避免介绍未定义的概念（例如描述性规范和命令性规范）---尽量使用一些一般性易理解的术语---尽管作者对于该领域较为熟悉，但并不意味着潜在的读者也知晓这些术语。与此同时，在结果方面的描述也不明确，不清楚本研究在哪些方面有新的发现。此外，也需要简要地概括本研究结果在哪些方面有可能的应用，或者至少是介绍研究结果的意义。为什么读者应该对这些结果产生一定的兴趣？总而言之，摘要应该重新再修改清楚。

修改说明 2：感谢专家的宝贵意见，我们根据您的建议重写了摘要，突出了本文的研究背景、主要发现和这些发现的理论意义，并改正了您所说的专业术语过多的问题。修改后的摘要如下：

第三方惩罚对合作的维系可能来自其经济功能，即惩罚降低了违规收益；但也有学者认为其规范激活功能同样重要。然而，由于先前研究没有区分惩罚的这两种功能，因而未能回答：当惩罚不足以影响违规收益时，是否还能有效促进合作？通过操作违规的不同收益，实验结果显示，即使违规收益大于被惩罚的损失，第三方惩罚依然能抑制自利行为并提升合作。这表明惩罚的规范激活功能并非以经济功能为前提。后两个实验进一步证实惩罚对合作的促进在很大程度上是通过规范激活来实现的，并且存在两种溢出效应：惩罚激活了社会规范，并抑制了曾经的违规者（纵向溢出效应）和旁观者（横向溢出效应）在新博弈情境下的自私行为，尽管在新情境下并不存在惩罚机制。对基于规范激活的溢出效应的发现补充了文献中占主导地位的经济学解释，并为理解人类社会长时间、大规模的合作提供了新视角。

意见 3：英文摘要：应该避免缩写（例如“don’t”）；第三段的首句话太长。

修改说明 3：感谢专家的指正，我们遵循您的意见对上述问题进行了修改。

参考文献

陈思静, 何铨, 马剑虹. (2015). 第三方惩罚对合作行为的影响: 基于社会规范激活的解释. *心理学报, 47(3)*, 389–405.

Bicchieri, C., Dimant, E., & Xiao, E. T. (2018). *Deviant or wrong? The effects of norm information on the efficacy of punishment* (PPE Working Papers 0016). Philadelphia, PA: Philosophy, Politics and Economics of University of Pennsylvania.

Cialdini, B., Kallgren, A., & Reno, R. (1991). A focus theory of normative conduct. *Advances in Experimental Social Psychology*, *24*, 201–234.

Fehr, E., & Williams, T. (2018). *Social norms, endogenous sorting and the culture of cooperation*. (ECON Working Papers 267). Zurich, Switzerland: Department of Economics of University of Zurich.

Gintis, H., & Fehr, E. (2012). The social structure of cooperation and punishment. Behavioral and Brain Sciences, 35(1), 28–29.

Lois, G., & Wessa, M. (2019). Creating sanctioning norms in the lab: The influence of descriptive norms in third-party punishment. *Social Influence*, *14*(2), 50–63.

……

专家 2 的审稿意见及修改说明

文章行文规范，思路清晰；实验设计完善、3 个实验之间的逻辑关系明确，具有一定的理论和实践意义，但是仍有一些问题需要改善。

意见 1：文章的三个理论贡献建议在引言部分和讨论部分进一步明确和呼应论证；

修改说明 1：感谢专家的宝贵意见，您的建议对本文大有裨益。我们遵循您的建议，重新撰写了前言、3 个实验的讨论部分以及总讨论，突出了本研究所关注的 4 个研究问题。在前言中，我们详细论证了是如何提出这 4 个问题的，而在实验讨论与总讨论中，我们又对这 4 个研究问题进行了回应，并探讨了我们的新发现在现有文献知识结构中的作用与意义，从而使整个文章的脉络更加清楚。修改部分较多，烦请专家参看前言、讨论与总讨论中标蓝文字。

意见 2：需要进一步阐述实验 2 中独裁者博弈中的（1）和（2）加权平均数为何可以代表两种社会规范的水平。即需补充这两种中介变量的操纵性定义。

修改说明 2：感谢专家的宝贵建议，这有利于我们更清楚地表达本文观点。遵循您的意见，我们在原文中补充了有关描述性和命令性规范激活水平的操作定义和文献来源。修改后文字如下（位于正文 3.2 部分第 2 段）：

我们用 1）和 2）项数字来计算被试在博弈中描述性和命令性规范的激活水平。现有文献中研究者常使用被试对某个行为（态度）在人群中普遍程度的百分比估计来表示其描述性（命令性）规范的激活水平（Voisin et al., 2016），如 Bicchieri 和 Xiao（2009）以及 Chen 等（2020）在独裁者博弈中，让被试估计分配（认为应该分配）不同金额（10, 20, … 50）的个体分别占多大比例，以此计算分配（认为应该分配）金额的加权平均值，并以此来代表被

试的描述性（命令性）规范的激活水平。和上述操作类似，在实验 2 中我们用 1）和 2）这两项各自的加权平均值分别作为描述性规范和命令性规范激活水平的操作定义。

意见 3：采用 PROCESS 分析时需具体说明使用哪一 Model 进行中介分析。

修改说明 3：感谢专家指出的问题，我们已在正文中补充说明了我们采用的是 PROCESS 插件中的 Model 4 进行中介效应检验。正文中修改如下（位于正文 3.3 部分第 2 段）：“因此我们使用 Preacher 和 Hayes（2004）所开发 PROCESS3.5 插件进行中介效应检验（Model 4）。”

意见 4：实验 2 中介结果呈现部分：严格意义上来讲，不能因为描述性规范中介的效应量大于命令性规范中介，就得出“描述性规范的作用更大”的结论，只有通过进一步差异检验才能证明。可以说描述性规范的效应量大于命令性规范。

修改说明 4：感谢专家的审慎阅读，您的严谨使我们受益良多。原文中这一表述确实不够严谨，针对该问题我们首先修改了正文中的表述，其次我们根据您的建议进一步检验了两种规范中介作用的差异，并将该结果报告在正文中。具体修改如下（位于正文 3.3 部分第 2 段）：

两种规范的间接效应占总效应的 77.20%，其中描述性规范的间接效应占 53.08%，命令性规范占 24.12%（图 3），并且两种规范间接效应的大小差异不显著（BootSE = 0.09，BootLLCI = -0.006，BootULCI = 0.304）。

意见 5：实验 3 中实验操纵是否改变了中介变量的水平，即描述性规范和命令性规范的水平？建议在做中介分析之前补充自变量影响中介变量的分析。

修改说明 5：感谢专家的中肯建议，我们针对这一问题首先补充说明了关于实验组与对照组两种规范差异的 *t* 检验结果，在正文中增加了对该结果的说明如下（位于正文 4.3 部分第 2 段）：

违规组被试的描述性规范（$M = 3.37$，$SD = 2.20$）显著高于规范组（$M = 2.98$，$SD = 1.89$）（$t = 2.30$，$p = 0.023$，$d = 0.36$，95%C.I. = [0.06，0.73]），违规组被试的命令性规范（$M = 4.97$，$SD = 2.77$）也显著高于规范组（$M = 4.18$，$SD = 2.51$）（$t = 3.32$，$p = 0.001$，$d = 0.52$，95%C.I. = [0.32，0.1.27]），这说明被试合作行为的提高有可能是由于旁观惩罚而激活了两种社会规范。

意见 6：实验 3 结果中描述性规范中介效用显著，而命令性规范中介效用不显著，与实验 2 结果不同。作者在这里的描述：“这一方面在一定程度上支持了实验 2 的结果，即惩罚通过激活人们的社会规范来提升他们的合作水平”是不恰当的。另外，也不能说“实验 2 和 3 的结果均显示描述性规范对合作的作用高于命令性规范”，因为命令性规范中介效用并不显著，不存在中介效用。作者应对为什么实验 3 中命令性规范中介效用不显著、以及实验 2 和实验 3 之间的差异进行讨论，建议补充。

修改说明 6：再一次感谢专家的宝贵意见，这确实提升了本文在表述上的严谨性。我们根据您的意见，修改了这部分的文字表述，并增加了对结果的讨论，包括为什么实验 3 中命令性规范的中介作用不显著，以及实验 2 和 3 之间的异同。修改文字如下（位于正文 4.3 部分第 3 段）：

比较描述性规范和命令性规范这两条路径，我们看到实验操作确实同时激活了这两种规范，对两种规范在实验操作前后的平均数差异检验也验证了这一点，两者间的差别主要体现在激活后的描述性规范提升了被试的合作水平，但命令性规范却未能起到类似作用。对上述结果的一种解释是在大部分情况下人们更容易受到描述性规范的影响（陈思静等，2015; Cialdini et al., 1991），因为描述性规范涉及的是事实判断（人们是怎么做的？），而命令性规范涉及价值判断（人们认为应该怎么做？），个体对事实判断的信息处理速度要高于对价值判断的处理（Deutsch & Gerard, 1955）。进一步比较实验 2 和 3 的结果，可以看到一个明显的差异是在实验 2 中描述性规范和命令性规范的中介效应均显著，且无显著差异，尽管单纯从数字上来看，前者的效应略高于后者，而在实验 3 中描述性规范的中介作用显著，命令性规范不显著，我们推测这可能是因为两个实验中被试的个人卷入度（personal involvement）有所不同：在实验 2 中，被试在第一阶段亲身经历了惩罚，而在实验 3 中被试仅仅旁观了他人受罚，因此可以合理地认为被试在实验 2 中的个人卷入度更高。Petty 和 Cacioppo（1986）指出，当个人卷入度较高时，命令性规范对行为的作用更为明显，这一观点可以解释实验 2 和 3 的差异：由于实验 2 中被试的卷入度更高，因此命令性规范对合作行为的作用也就更为明显，而在实验 3 中低个人卷入度导致命令性规范的影响不显著。

参考文献

陈思静, 何铨, 马剑虹. (2015). 第三方惩罚对合作行为的影响: 基于社会规范激活的解释. *心理学报, 47(3)*, 389–405.

Cialdini, B., Kallgren, A., & Reno, R. (1991). A focus theory of normative conduct. *Advances in Experimental Social Psychology, 24*, 201–234.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629-636.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York, NY: Springer

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 717–731.

Voisin, D., Girandola, F., David, M., & Aim, M. A. (2016). Self-affirmation and an incongruent drinking norm: Alcohol abuse prevention messages targeting young people. *Self and Identity*, *15*(3), 262–282.

……..

专家 3 的审稿意见及修改说明

修改意见：总体来说，该研究通过三个实证研究，并且三个研究都采集了较大的样本量，论证了第三方惩罚作为规范提示作用能够促进非亲缘关系间的群体合作。基于社会规范的角度也为研究第三方惩罚是如何改变人们的行为提供了新的视角。

意见 1：摘要在表述完相应的研究结果之后，应该再用一句话总结出文章所得到的更深层次的结论。

修改说明 1：感谢专家的宝贵意见，我们重新撰写了摘要，并根据您的建议在摘要最后增加了总结性的讨论。修改后的摘要如下：

第三方惩罚对合作的维系可能来自其经济功能，即惩罚降低了违规收益；但也有学者认为其规范激活功能同样重要。然而，由于先前研究没有区分惩罚的这两种功能，因而未能回答：当惩罚不足以影响违规收益时，是否还能有效促进合作？通过操作违规的不同收益，实验结果显示，即使违规收益大于被惩罚的损失，第三方惩罚依然能抑制自利行为并提升合作。这表明惩罚的规范激活功能并非以经济功能为前提。后两个实验进一步证实惩罚对合作的促进在很大程度上是通过规范激活来实现的，并且存在两种溢出效应：惩罚激活了社会规范，并抑制了曾经的违规者（纵向溢出效应）和旁观者（横向溢出效应）在新博弈情境下的自私行为，尽管在新情境下并不存在惩罚机制。对基于规范激活的溢出效应的发现补充了文献中占主导地位的经济学解释，并为理解人类社会长时间、大规模的合作提供了新视角。

意见 2：在引言当中，在提出纯粹经济人理论无法完全解释第三方惩罚对违规作用的抑制作用是，提到一个研究是“当惩罚动机具有明显的自利特征时，惩罚不仅无法有效抑制违规行为，反而导致合作水平的下降。”这个地方表述不大明朗，可以再清楚阐述惩罚动机的自利

特征，以及这种动机如何改变了违规者的收益结构但却没有促进合作。这样更好能与下文进行链接。

修改说明 2：感谢专家富有建设性意见，我们重新阅读了这部分文字，发现我们的说明确实不够清楚，所以我们对这一部分进行了重新表述，并举例说明了受罚者对惩罚动机的感知是如何影响合作的。修改后文字如下（位于正文前言部分第 2 段）：

先前有研究者发现惩罚者的动机显著影响了惩罚的作用（谢东杰, 苏彦捷, 2019; Raihani & Bshary, 2015），如 Rand, Dreber, Ellingsen, Fudenberg 和 Nowak（2009）指出，惩罚是否被认为合理可以极大地影响受罚者的反应；而 Fehr 和 Rockenbach（2003）也注意到，当惩罚被认为是出于恶意（比如惩罚是为了获取更多的个人利益），尽管惩罚能显著降低违规收益（减少的金额等于初始金额的 40%），但受罚的违规者并没有表现出更高的合作水平，结果恰恰相反，其合作水平明显下降了。如果惩罚促进合作主要是由于其降低了违规收益，那么上述发现便难以得到合理的解释。

意见 3：引言部分分别提出了三个研究目的。是否可以增加一段总结的段落，言简意赅的阐明这三个研究之间内在联系。

修改说明 3：感谢专家的中肯意见，我们按照您的建议，在前言部分补充了一个段落，从总体上概括了本研究的主要内容和内在逻辑。补充的文字如下（位于正文前言部分第 5 段）：

概括而言，本文拟从社会规范的视角来解释第三方惩罚对合作的影响机制：我们认为规范激活是第三方惩罚的一种独立功能，即便无法降低违规收益，第三方惩罚依然可以抑制（促进）个体的违规（合作）行为（实验 1），同时，这一效应还溢出到了缺乏惩罚机制的新场景中（实验 2）和目睹惩罚行为的旁观者上（实验 3）。此外，我们还检验了两种规范在上述过程中的作用机制（实验 2 和 3），并讨论了新发现的理论和实践意义。

意见 4：关于实验二，用被试估计的“从 0-10 选择一个整数代表自己愿意分配给接受者的金额”来作为被试在独裁者博弈中被试的合作水平，对此表示有点疑问，这里的独裁者任务似乎并没有涉及到合作，因为接受方并没有自主权力。更多的可能体现的被试的公平性。作者应该解释为什么可以利用独裁者博弈任务来反应被试的合作。

意见 7：关于实验三，类似的，让被试“假设自己为分配者，从 0-10 选择一个整数代表自己愿意分配给接受者的金额”何以能够代表被试合作水平？请阐述。

修改说明 4 和 7：由于专家所提出的第 4 和第 7 点意见都集中在同一个方面，因此我们对上述意见统一进行说明。首先感谢专家的宝贵意见，您的建议使我们受益颇多。对上述问题的忽略确实是我们的疏漏之处，我们希望在这里做一点补充说明。合作（cooperation）一词在社会科学和生物学中具有丰富多彩的定义，而《科学》杂志尽管在 2005 年提出了包括合作在内的 25 个决定未来研究方向的重大问题（Kennedy & Norman, 2005），但并没有对合作下过一个明确的定义。不过有关合作，一个较为常见的定义是“个体付出成本以使他人受益的行为”（e.g., 韦倩等, 2019; Nowak, 2006; Rand, 2016）。黄少安和张苏（2013）在考察了现有文献中有关合作的多个定义后也指出“合作是自己付出成本 c，向其他人或者公共品提供价值为 b 的贡献的行为”。这个定义包括了本文在两种实验范式下被试的行为：在公共物品博弈中（实验 1），合作表现为个体付出代币从而增加资金池（公共物品）的行为，而在独裁者博弈中（实验 2 和 3），合作表现为独裁者向接受者分配金额而使后者受益的行为。现有文献也有大量的例子可以佐证上述观点，如 Haselhuhn 和 Mellers（2005）以及 Raihani 和 Bshary（2012）都将独裁者博弈中独裁者分配金额的行为称之为 cooperate, cooperation 或 cooperative behavior。基于上述文献，我们认为用独裁者的分配金额来测量被试的合作水平还是符合学界规范的。为了更清楚地体现我们的观点，我们遵循专家的意见，在原文中补充了有关合作的定义，从而方便读者理解合作在本文中合作的准确含义。补充文字如下（位于正文 3.2 部分第 2 段）：

根据黄少安和张苏（2013）对合作所下定义：合作是自己付出成本而使其他人或者公共物品受益的行为，我们用上述 3）和 4）项数字分别表示被试在两种博弈情形下的合作水平（在独裁者博弈中，合作意味着使对方受益；而在公共物品博弈总，合作意味着自己的行为提高了公共物品的产出），数字越大表示合作水平越高。

意见 5：实验二的研究结果中提到了使用 Bootstrap 的方法进行中介效应检验。这与一般的中介效应检验方法有所不同，那么选用这个方法的依据和理由是什么？这种检验手段的优势在哪里？作者应对此进行解释。

修改说明 5：感谢专家的宝贵意见，我们在中介效应检验的结果报告前增加了对选用 Bootstrap 方法的简要说明（位于正文 3.3 部分第 2 段）：

需要说明的是，有研究者指出用偏差校正的非参数百分位 Bootstrap 法计算系数乘积的置信区间比 Sobel 法得到的置信区间更精确（方杰, 张敏强, 2012; 温忠麟, 叶宝娟, 2014），

因此我们使用 Preacher 和 Hayes（2004）所开发 PROCESS3.5 插件进行中介效应检验（Model 4）。

意见 6：关于对实验二结果的讨论，前面提到“人们内化了这种合作规范并直觉性地将之应用到不同的情境中去”，后面又说在新的情境下会引发“有意识的理性思考“，这里有矛盾冲突之意，作者这里应该再仔细阐述一下，这里是指在相似的情形下，引发的是直觉性反应，而在相异的情形下才引发的理性思考吗？

修改说明 6：感谢专家的宝贵意见，您的严谨极大地提高了本文的写作质量。在原文中，根据 Rand 等（2014）所提出的社会启发法假说，我们提出“人们内化了这种合作规范并直觉性地将之应用到不同的情境中去”，这里的“不同的情境”我们本意是指“各种情境”而不是“不相同的情境”，也就是说，我们希望表达的意思是：人们内化了合作规范并将这种规范运用在了各类场景中，但随着场景的相异程度提高，人们有意识思考的程度也会有所增加，而这种有意识的思考恰好阻碍了人们的合作行为。但正如专家所言，由于我们的疏忽，这部分表述确实存在让人误解的地方。根据您的建议，我们对这一部分的文字进行了重新表述，修改后文字如下（位于正文 3.3 部分第 4 段）：

上述结果一方面进一步证实了惩罚的溢出效应，另一方面也意味着惩罚通过激活社会规范所带来的合作提升效果虽然可以跨情境迁移，但不同情境下提升效果比相同情境低。这一结果可以通过 Rand 等（2014）所提出的社会启发法假说（social heuristics hypothesis）得到解释：真实生活中个体间的互动往往是非匿名的和重复博弈的（Dreber et al., 2008; Rand et al., 2016），从长远来看合作是更有利的博弈策略，长此以往，人们内化了这种合作规范并直觉性地将之应用到各种情境中去，但新情境的不同会激发个体的有意识思考，而通过这种思考人们会发现对自身利益而言在新的情境中合作未必是最佳选择（Peysakhovich & Rand, 2016），换言之，理性思考会抑制个体在新情境中的合作行为。

意见 8：关于总讨论，作者的分析和讨论有一定的深度，把研究结果和现实意义紧密结合。有一点建议是如果作者在研究意义的第一段，提出“惩罚的道德合法性是惩罚发挥积极做作用的必要条件”的阐述之后，能够结合自己的研究，说明该研究当中的惩罚是如何符合社会规范等的讨论，也能更好的阐明该研究的发现如何更能解释前任的研究发现。

修改说明 8：感谢专家对我们的肯定，同时专家的建议也让我们受益匪浅，我们根据您的意见，联系以往文献补充了有关论述，从而使我们的讨论更加丰满。补充文字如下（位于正文 5.1 部分第 2 段）：

上述观点的一个推论是如果惩罚完全不具备经济功能，那么我们可以在很大程度上排除惩罚的不合理动机（如惩罚是为了提高自身的相对优势），在这种情况下，按照实验 1 的结果，我们应该能观察到这类惩罚对合作同样具有促进作用。事实上，确实有研究者注意到，面对违规行为，他人的言语责备（也有学者将言语责备称为社会惩罚或道德惩罚）就能起到类似的作用（Noussair & Tucker, 2005），而无需对违规者造成具体的金钱或物质损失，甚至比以降低经济收益为目标的惩罚效果更好（Wu, Balliet, & van Lange, 2016）。实验 1 的结果可以解释上述现象：尽管言语责备并未改变惩罚的收益，但和第三方惩罚类似，言语责备起到了提示违规者存在某种规范的作用，同时言语责备在很大程度上排除了为自身牟利的非法动机，从而有效地降低了违规者的自私行为。当然，也有研究者认为言语责备的作用同样可以从经济角度来解释，比如 van den Berg, Molleman 和 Weissing（2012）认为成本表现为多种形式，言语责备尽管未必会提高违规行为的金钱成本，但可能提高了违规者在人际关系方面的成本，因此实际上依然减少了违规者的收益。然而，实验室环境中的言语责备往往程度较轻，如“我认为你的分配方案不公平”（Nelissen & Mulder, 2013）或“某某人只关心自己”等（崔丽莹等, 2017），且经常发生在匿名环境中（陈思静, 徐烨超, 2020），因而似乎很难认为匿名状态下上述言语能对违规者的人际利益造成实质性损害。综上所述，我们认为实验 1 的结果可以更好地解释言语责备对合作的提升作用。

意见 9：关于表 2，表 2 的内容中关于有两个变量，M1 和 M2，都是合作行为。则这两个因变量有什么区别吗？文中没有明确解释，以至于难以完全理解实验 2 的中介效应检验的结果。

修改说明 9：感谢专家指出的问题，文中对于中介作用的说明确实缺少了自变量到两个中介变量的回归分析结果，针对这一问题我们在中介检验结果的表格中补充了自变量（惩罚）对两个中介变量（描述性规范和命令性规范）的回归分析结果。结果如表 2 所示，并在正文中增加了对该结果的说明（位于正文 3.3 部分第 2 段）：

检验结果如表 2 所示：M_1 和 M_2 中惩罚对两种规范都有显著的影响。与 M_3 相比，M_4 在引入两种规范后 R^2 增加了 0.24，意味着引入两种规范能解释合作行为变异的 24%。

表 2 中介效应的检验

变量	M_1（描述性规范）		M_2（命令性规范）		M_3（合作行为）		M_4（合作行为）	
	系数	***SE***	系数	***SE***	系数	***SE***	系数	***SE***
常数	1.87^{***}	0.51	2.57^{***}	0.63	1.38^{*}	0.66	-0.17	0.62
惩罚	0.97^{**}	0.32	1.53^{***}	0.40	1.08^{**}	0.42	0.25	0.39
描述性规范							0.59^{***}	0.08
命令性规范							0.17^{*}	0.07
模型	R^2	***MSE***	R^2	***MSE***	R^2	***MSE***	R^2	***MSE***
	0.048	4.72	0.076	7.17	0.037	2.79	0.278	5.905

注：括号内为因变量，$^{***}p < 0.001$，$^{**}p < 0.01$，$^{*}p < 0.05$。

参考文献

陈思静, 徐烨超. (2020). “仁者”还是“智者”: 第三方惩罚对惩罚者声誉的影响. *心理学报*, *52*(12), 1436–1451.

崔丽莹, 何幸, 罗俊龙, 黄晓娇, 曹玮佳, 陈晓梅. (2017). 道德与关系惩罚对初中生公共物品困境中合作行为的影响. *心理学报*, *49*(10), 1322–1333.

方杰, 张敏强. (2012). 中介效应的点估计和区间估计: 乘积分布法、非参数 Bootstrap 和 MCMC 法. *心理学报*, *44*(10), 1408–1420.

黄少安, 张苏. (2013). 人类的合作及其演进: 研究综述和评论. *中国社会科学*, *7*, 79–91.

韦倩, 孙瑞琪, 姜树广, 叶航. (2019). 协调性惩罚与人类合作的演化. *经济研究*, (7), 174–187.

温忠麟, 叶宝娟. (2014) 有调节的中介模型检验方法: 竞争还是替补?. *心理学报*, *46*(5), 714–726.

谢东杰, 苏彦捷. (2019). 第三方惩罚的演化与认知机制. *心理科学*, *42*(1), 216–222.

Dreber, A., Rand, D. G., Fudenberg, D., & Nowak, M. A. (2008). Winners don’t punish. *Nature*, *452*(7185), 348–351.

Fehr, E., & Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature*, *422*(6928), 137–140.

Haselhuhn, M. P., & Mellers, B. A. (2005). Emotions and cooperation in economic games. *Cognitive Brain Research*, *23*(1), 24–33.

Kennedy, D., & Norman, C. (2005). What don’t we know. *Science*, *309*(5731), 75–77.

Nelissen, R. M., & Mulder, L. B. (2013). What makes a sanction “stick”? The effects of financial and social sanctions on norm compliance. *Social Influence*, *8*(1), 70–80.

Noussair, C., & Tucker, S. (2005). Combining monetary and social sanctions to promote cooperation. *Economic Inquiry*, *3*(3), 649–660.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*(5805), 1560–1563.

Peysakhovich, A., & Rand, D. G. (2016). Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science*, *62*(3), 631–647.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 717–731.

Raihani, N. J., & Bshary, R. (2012). A positive effect of flowers rather than eye images in a large-scale, cross-cultural dictator game. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1742), 3556–3564.

Raihani, N. J., & Bshary, R. (2015). The reputation of punishers. *Trends in Ecology and Evolution*, *30*(2), 98–103.

Rand, D. G. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science*, *27*(9), 1192–1206.

Rand, D. G., Brescoll, V. L., Everett, J. A., Capraro, V., & Barcelo, H. (2016). Social heuristics and social roles: Intuition favors altruism for women but not for men. *Journal of Experimental Psychology: General*, *145*(4), 389–396.

Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science*, *325*(5945), 1272–1275.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, 3677.

van den Berg, P., Molleman, L., & Weissing, F. J. (2012). The social costs of punishment. *Behavioral and Brain Sciences*, *35*(1), 42–43.

Wu, J., Balliet, D., & van Lange, P. A. (2016). Gossip versus punishment: The efficiency of reputation to promote and maintain cooperation. *Scientific Reports*, *6*, 23919.

第二轮

专家 1 的审稿意见及修改说明

文章对理论的交代和结果的分析教之前有了进一步地改善。但请作者再进一步就以下两个问题进行解释：

意见 1：中文摘要：有关研究目的的描述比以前清楚了，但有关实验设计及其结果的叙述显得太过简略。能否按照常规的心理学论文摘要，描述清楚比如被试数量、不同实验操纵在结果上的不同等一些必要的信息。

修改说明 1：感谢专家富有建设性的意见，我们已经遵循您的建议重写了摘要，新摘要如下：

当惩罚不足以影响违规收益时，是否还能有效促进合作？实验一（N = 252）操作了违规的不同收益，相比于对照组，尽管存在第三方惩罚，高收益组的违规收益仍较高。然而，高收益组的合作水平显著高于对照组，这说明即使第三方惩罚无法降低违规收益，依然能抑制自利行为。实验二（N = 179）中相较于在第一阶段中未受惩罚的违规者，受罚的违规者在新博弈中表现出了更高的合作水平。2（是否旁观惩罚）×2（旁观前后）设计的实验三（N = 179）结果显示，旁观惩罚后被试的合作水平显著高于其它几种条件。在实验二、三中，社会规范在惩罚与合作间均起中介作用。这进一步证实惩罚对合作的促进在很大程度上是通过规范激活来实现的，并且存在两种溢出效应：惩罚激活了社会规范，并抑制了曾经的违规者（纵向溢出效应）和旁观者（横向溢出效应）在新博弈情境下的自私行为，尽管在新情境

下并不存在惩罚机制。对基于规范激活的溢出效应的发现补充了文献中占主导地位的经济学解释，并为理解人类社会长时间、大规模的合作提供了新视角。

意见 2：描述性和命令性社会规范实验操纵的有效性：我同意专家 2 提出的，有关作为中介变量的这两种规范的操纵性定义的意见，亦即分别以被试“估计分配”和“认为应该分配”的金额的加权平均代表这两种规范，缺乏充分的依据。虽然文章增加了 Bicchieri & Xiao (2009)、Voisin et al. (2016)和 Chen et al. (2020)文献试图作为该操纵的依据，然而前两个文献并没有如文章所言“让被试估计分配（认为应该分配）不同金额（10, 20, … 50）的个体分别占多大比例，以此计算分配（认为应该分配）金额的加权平均值，并以此来代表被试的描述性（命令性）规范的激活水平”（p. 23）。

修改说明 2：感谢专家的宝贵意见，您的严谨使我们受益匪浅。我们重新阅读了我们所引用的文献，确实存在不够严谨的地方。我们按照您的意见对这部分文字进行了重写，现说明如下。首先，就本文作者所知，关于社会规范感知的测量，目前在社会心理学文献中有三种比较常见的方式：1）被试估计某一人群中实施（赞同实施）某一行为的人数比例，如 Bicchieri & Xiao（2009）、Clapp 等（2003）、Sood 等（2020）以及 Voisin 等（2016）均采用了这种方式去估计被试的社会规范激活水平；2）采用 Likert 量表形式让被试对有关行为普遍性的陈述做出“完全不同意”到“完全同意”的评估，典型的表述如“我周围大多数人在流感期间积极采取了防护行为”，Liao 等（2019）和 Sparks 等（2014）的研究均属于这一类；3）第三种也正是本文所采用的方式相对较为少见，我们主要参考了 Chen 等（2020）的方法，这种方法同时考虑了比例与权重，信息的利用也较充分，因此我们在考虑因变量的操作定义时主要参考了这种方法。但正如专家指出，这种方式采用的研究并不多见，因此方法的有效性似乎还需更多后续研究的支持。为了弥补上述缺陷，我们借鉴了经济学领域中常见的一种稳健性分析方法，即采用结果变量的另一种操作定义来检验不同测量方式下结论是否有质的差异（e. g., 杜勇等, 2019; 孔东民等, 2017; Crouzet & Mehrotra, 2020），如果两种方法得出了相似的结果，那么就可以在一定程度上说我们的结论具有较高的稳健性。

具体而言，我们在修改稿中重新引入了上述测量社会规范的第一种也是最普遍的方式，即让被试估计某个行为的普遍程度。由于本研究主要考察有关合作的社会规范，因此我们在实验中要求被试估计做出了合作行为（将 7、8、9 或 10 代币分配给接受者）的分配者的百分比，以此作为规范激活水平的第二种操作定义，并采用这种操作定义对文章中原有结论进行了稳健性检验。结果显示，无论采用第一种操作定义（加权平均）还是第二种（百分比估

计），我们都得到了相似的结论。增加的文字主要展现在正文 3.2 部分第 3 段、3.3 部分第 2 段和第 4 段、4.2 部分第 2 段、4.3 部分第 4 段中标绿文字。由于修改部分较多且较为分散，不便在这里赘述，烦请专家移步正文审阅。

最后，我们补充说明下我们采用何种标准来判断某个分配方案是否为合作行为。先前有相当文献表明在不同文化语境中人们对于什么样的分配方案算是违规/合作有高度稳定的看法，即分配给对方的金额约小于 30%是一种违规行为（Csukly et al., 2011, Fehr & Fischbacher, 2003），且有学者认为这种在划分标准上的稳定性具有一定的生物学基础（Wallace et al., 2007）。以初始金额（20 代币）30%计算，6 代币为分界点，也就是说高于 6 代币的分配方案可被认为是一种合作行为，因此我们选择分配 7、8、9 和 10 代币作为合作行为的操作定义。

再次感谢您的意见，这对提升本文写作质量有极大的帮助！

参考文献

杜勇，谢瑾，陈建英. (2019). CEO 金融背景与实体企业金融化. *中国工业经济*, (5), 136–154.

孔东民，徐茗丽，孔高文. (2017). 企业内部薪酬差距与创新. *经济研究*, *10*, 144–157.

Bicchieri, C., & Xiao, E. (2009). Do the right thing: But only if others do so. *Journal of Behavioral Decision Making*, *22*(2), 191–208.

Chen, H., Zeng, Z., & Ma, J. (2020). The source of punishment matters: Third-party punishment restrains observers from selfish behaviors better than does second-party punishment by shaping norm perceptions. *PloS One*, *15*(3), e0229510.

Clapp, J. D., Lange, J. E., Russell, C., Shillington, A., & Voas, R. B. (2003). A failed norms social marketing campaign. *Journal of Studies on Alcohol*, *64*(3), 409–414.

Crouzet, N., & Mehrotra, N. (2020). Small and large firms over the business cycle. *American Economic Review*, *110* (11), 3549–3601.

Csukly, G., Polgár, P., Tombor, L., Réthelyi, J., & Kéri, S. (2011). Are patients with schizophrenia rational maximizers? Evidence from an ultimatum game study. *Psychiatry Research*, *187*(1–2), 11–17.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–791.

Liao, Q., Wu, P., Wing Tak Lam, W., Cowling, B. J., & Fielding, R. (2019). Trajectories of public psycho-behavioural responses relating to influenza A (H7N9) over the winter of 2014-15 in Hong Kong. *Psychology & Health*, *34*(2), 162–180.

Sood, S., Kostizak, K., Lapsansky, C., Cronin, C., Stevens, S., Jubero, M., ... & Obregon, R. (2020). ACT: An evidence-based macro framework to examine how communication approaches can change social norms around Female Genital Mutilation. *Frontiers in Communication*, *5*, 29.

Sparks, P., Hinds, J., Curnock, S., & Pavey, L. (2014). Connectedness and its consequences: a study of relationships with the natural environment. *Journal of Applied Social Psychology*, *44*(3), 166–174.

Voisin, D., Girandola, F., David, M., & Aim, M. A. (2016). Self-affirmation and an incongruent drinking norm: Alcohol abuse prevention messages targeting young people. *Self and Identity*, *15*(3), 262–282.

Wallace, B., Cesarini, D., Lichtenstein, P., & Johannesson, M. (2007). Heritability of ultimatum game responder behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(40), 15631–15634.

……

专家 2 的审稿意见及修改说明

经过第一轮修改后，文章在研究问题引出、整体逻辑和结果讨论部分均有提高，但仍存在部分问题需要修改。

意见 1：引言部分，建议将 4 个研究问题单独列出，不要放在段落中；或者以实验假设的方式单独列出，让读者能够快速了解本研究的科学问题。

修改说明 1：感谢专家的宝贵意见，我们遵循您的建议，对前言的部分文字进行了重新表述，并把本文的 4 个主要研究问题单独列出，以方便读者快速了解本研究的主要关注点。修改部分在正文前言部分以标绿文字展示。

意见 2：实验 1：对于实验实施过程的阐述不清晰。如投入和收益部分：“投入 10 代币（合作）只能给自己带来 5 代币的收益，净损失为 5 代币”，但这个情况只发生在别人都不投入的情况下，而不是在其他人投入的情况下发生，文章阐述不清晰可能导致读者混淆。此外，对于实验组设置，在实验设计和过程中为：3（对照组、高收益组和低收益组）；结果部分却阐述为“低成本组”、“高成本组”和控制组（图 1）。是怎么样的对应关系？请作者修改。

修改说明 2：感谢专家的细心阅读与及时提醒。这确实是我们的一个疏忽，在第一轮修改中由于前后表述不统一而造成了混淆。我们在这里简单说明下：在本研究中，高成本意味着违规者需要为自己的违规行为支付较高的成本，因而违规收益较低，所以“高成本组”和“低收益组”是同一个东西，“低成本组”和“高收益组”同理。但诚如专家所言，前后表述不一可能会给读者造成混淆，因此我们按照您的建议对相关部分都进行了修改，统一表述为“高收益组”和“低收益组”。

意见 3：对于实验 1 中的“第三方惩罚”实验操纵存在疑问。根据先前文献：“第三方惩罚中实施惩罚的被试是不参与博弈的第三方, 其利益不相关的，因此这种惩罚并不是出于个人利益的考虑(Fehr & G ächter, 2002)”。而本文实验 1 中第三方的设置为：参加博弈的被试可以惩罚没有遵守规则的成员，即惩罚者本身就是利益获得者，其他成员的不投入会直接导致其收益减少。因此可能存在对不遵守规则成员的报复性惩罚，实验操纵与第三方惩罚的概念不符，请作者斟酌。

修改说明 3：感谢专家的宝贵建议，这确实是我们的一个疏忽，由于本文重点在于研究 2 和 3，因此在写作时我们主要在关注后两个实验中被试的行为，导致我们忽视了实验 1 中的设置，犯下了概念混淆的错误。您的建议对我们是一个及时的警醒，为了弥补这个疏漏，我们在 12 月初收到您的意见反馈后，立刻招募了新被试重做了实验 1。在新的实验 1 中，被试被分成了两类。参与者：这类被试参与公共物品博弈；执行者：这类被试不参与公共物品博弈，但可选择对违规者实施惩罚，惩罚成本由执行者承担，并直接和其实验报酬挂钩。这一设置使得新实验中的惩罚成为了一种明确的第三方惩罚，因为违规行为并不影响执行者的收益。由于我们重写了整个研究 1，修改后文字较多，不便在这里赘述，烦请专家移步正文重新审阅。再次感谢您的费心阅读和宝贵建议，这对我们的研究大有助益！

意见 4：在写作上: 文章存在大段陈诉的现象，部分段落冗长。如实验设计部分，建议可适当分段并增加小标题，明确每个测量过程的目的，使得实验操纵更加清晰明确。同时一些较长的操纵依据可作为脚注，避免正文拖沓（如 p.23 两种规范选取的依据）。

修改说明 4：感谢专家的建议，我们已按照您的建议对有关段落进行了重新表述，必要的地方增加了小标题，从而使文字更为清晰简洁，并在适当的地方增加了若干脚注。由于修改部分较为分散，为了避免啰嗦，我们不再这里一一罗列，烦请专家移步正文审阅。

……

专家 3 的审稿意见及修改说明

作者对于所给出的修改意见都逐一做出了认真和详尽的修改，在根据意见修改过后，本研究的行文逻辑和研究思路都更加清晰。同意该文章进行发表。

回复：感谢专家对本文的肯定，这对我们是一个巨大的鼓舞!

第三轮

专家 1 的审稿意见及修改说明

相比三个实验和总讨论部分，文章在摘要和前言部分的写作上上下文逻辑和概念交代的不清晰。具体有以下几点：

意见 1：中文摘要的表述还是不清楚：首先，在阐述三个实验之前，作者仅用“当惩罚不足以影响违规收益时，是否还能有效促进合作？”作为理论铺垫的介绍，显得非常不充分。其次，没有交代清楚（1）实验一中的高收益组和对照组之间在收益方面的区别；（2）实验二

的有关设计（例如“第一阶段”所指为何、什么原因造成受罚的违规者表现出更高的合作水平、以什么来衡量合作水平等）；（3）实验三的其他几种条件指什么？（4）所谓“规范激活”需要具备什么条件，以及惩罚具体激活了什么“社会规范”？总之，摘要需要表达简明，避免笼统及含糊。

修改说明 1：感谢专家审慎的阅读，您的严谨细致让我们受益良多！原文摘要的确存在上述问题。我们根据您的意见逐条核对并修改了摘要，力求清晰地呈现出研究的主要结论。修改后的摘要如下：

第三方惩罚对合作的维系可能来自其经济功能，即惩罚降低了违规收益；但也有学者认为其规范提示功能同样重要——通过唤起个体的社会规范来提升合作。然而，由于先前研究没有区分惩罚的这两种功能，因而未能回答：当惩罚不足以影响违规收益时，是否还能有效促进合作？实验一（N = 252）操作了违规的不同收益，高收益组被试选择违规的预期收益总是大于等于无第三方的对照组，但高收益组的合作水平却显著高于对照组，这说明即使第三方惩罚无法降低违规收益，依然能抑制自利行为。实验二（N = 179）中相较于在第一阶段有第三方的独裁者博弈中未受惩罚的违规者，受罚的违规者在第二阶段无第三方的独裁者博弈中分配给对方的代币和公共物品博弈中投入公共账户的代币均表现出了更高的水平。中介模型检验显示，惩罚通过提升描述性和命令性规范有效促进了被试的合作水平。2（是否旁观惩罚）×2（旁观前后）设计的实验三（N = 179）显示，旁观惩罚后被试的合作水平显著高于旁观前，也高于未旁观惩罚的被试。与实验二类似，社会规范在惩罚与合作间也有显著的中介作用。这进一步证实惩罚对合作的促进在很大程度上是通过规范激活来实现的，并且存在两种溢出效应：按照社会规范聚焦理论，惩罚使合作的社会规范成为了被试当前的“焦点”，并抑制了曾经的违规者（纵向溢出效应）和旁观者（横向溢出效应）在新博弈情境下的自私行为，尽管在新情境下并不存在惩罚机制。这两种溢出效应的发现补充了文献中占主导地位的经济学解释，并为理解人类社会长时间、大规模的合作提供了新视角。

意见 2：前言部分第一段提到了“大量学者探讨了第三方惩罚与规范之间的关系”和“而我们注意到，在回答这个问题上，基于规范视角的研究是相对缺席的”，这两个表述似乎是自相矛盾。此外，这里所指的“规范”的用词太模糊。

修改说明 2：感谢专家的细心阅读和宝贵建议。我们在表述上的疏忽确实给读者造成一种矛盾的感觉，我们在这里做一简要解释。先前确实有相当数量的研究探讨了社会规范与惩罚的关系，但大部分着眼于社会规范在第三方惩罚促进合作中的调节作用，如 Bicchieri 等（2018）、

Fehr 和 Williams（2018）以及 Lois 和 Wessa（2019）均属于这一类型的研究，分别考察了社会规范的存在与否或社会规范的不同方式对惩罚-合作关系的影响；而在后面句子中，我们希望从另一个角度来探讨社会规范在惩罚-合作中的作用：我们认为，惩罚促进合作主要通过两个途径：1）改变违规行为的收益结构；2）激活违规者有关合作的社会规范。换言之，我们更多的是从中介作用的角度来探讨社会规范的作用。在这方面，基于规范视角的研究确实相对少见。陈思静等（2015）从社会规范激活的角度来理解第三方惩罚，从而为本文的研究思路提供了理论启发，但由于该研究未能分离惩罚两种作用：改变收益结构和激活社会规范，因此，本文希望在控制违规者收益结构的基础上将我们对社会规范激活在惩罚-合作关系中中介作用的理解推进一步。其次，您提出“规范”一词过于模糊，我们同意您的说法，并作了如下修改：1）补充了社会规范的有关定义；2）补充说明了在本文中规范指的是有关合作的社会规范，但为了避免行文啰嗦，我们在有些地方就直接表述为规范。我们希望通过上述修改能使读者更清晰地了解本文内容。有关专家的上述两点意见，我们在正文中做了相应修改，但由于修改内容较为分散，不便在这里赘述，烦请专家移步正文“前言”部分第 1 段和第 3 段中标橙文字。

意见 3：前言部分第三段提到了实验 1 的目的是“考察惩罚的规范效应”，以期为社会规范聚焦理论解释第三方惩罚提供证据。这里没有交代清楚什么是惩罚的规范效应，以及两者之间的关系。

修改说明 3：感谢专家的意见，您的严谨极大地提高了本文的写作质量！我们根据您的建议，在正文中简要地补充说明了惩罚的规范效应，同时也阐述了检验规范效应与验证社会规范聚焦理论之间的关系。修改后文字如下（位于正文前言第 3 段）：

然而，在先前研究中第三方惩罚通常改变了违规者的收益结构，这意味着先前研究者未能严格区分第三方惩罚的两种功能：通过降低违规收益来提升合作（惩罚的经济效应）以及通过激活社会规范来提升合作（惩罚的规范效应）。本文拟在这方面为现有文献提供有益补充，具体而言，本文将于实验 1 中在控制违规者收益的情况下检验第三方惩罚的规范激活功能。如果实验结果显示，尽管惩罚并未降低违规者的收益，但受罚的违规者依然表现出了较高的合作行为，那么我们就可以在一定程度上认为，惩罚的规范效应是一种独立于经济效应的功能，并且为社会规范聚焦理论提供了新的实证证据：激活人们的规范就可改变其行为。

意见 4：前言部分仅用一句话介绍描述性规范和命令性规范，缺少有关这两种规范与第三方惩罚研究的必要的文献回顾。因而，不清楚文章紧接着提出的，需要探索的研究问题 4 的合理性。

修改说明 4：感谢专家富有建设性的意见，我们遵循您的建议，重新写了这一部分，较为详细地解释了两种社会规范的定义和区别，从而更为清楚地引出本文的研究问题 4。修改后文字如下（位于正文前言第 5 段）

最后，社会规范作为被群体成员广泛接受并区别于法律规章的行为准则（Cialdini & Trost, 1998; Forquesato, 2016），在社会科学文献中通常被区分为描述性规范（descriptive norm）和命令性规范（injunctive norm）（Cialdini et al., 1991）：前者指的是人们在某一方面的普遍行为模式，如合作的描述性规范可理解为人们所表现出来的合作行为的普遍程度；而后者指的是人们对某一行为普遍所持赞成或批评的态度，如合作的命令性规范可理解为人们对他人合作行为的赞成程度。社会规范可显著影响人们的行为，如简化个体的行为决策并在个体面对复杂、不确定甚至是危险的情境时得到行为上的指引（McDonald & Crandall, 2015）。但需要说明的是，研究者从不同角度指出了两种规范在影响行为中的区别，如 Deutsch 和 Gerard（1955）指出人们对描述性规范的认知加工速度要高于对命令性规范的加工，因此，描述性规范通常更容易对行为产生影响；而 Petty 和 Cacioppo（1986）从个人卷入度（personal involvement）比较了两种规范对行为的影响，并指出当个人卷入度较高时，命令性规范的作用更大。就本文而言，一个值得探讨的问题是当惩罚通过激活社会规范来影响合作时，惩罚是激活了其中一种规范还是两种规范都有所激活？如果两种规范都被激活了，那么它们是否具有不同的作用机制？我们将在实验 2 和 3 中详细探讨这些问题。此外，由于社会规范聚焦理论的重点考察对象是描述性规范，如果我们的实验结果表明，在惩罚通过激活规范而影响合作的过程中，命令性规范也被激活并产生了显著影响，那么本文的结果也可在一定程度上被视为对这一理论的有益补充。

意见 5：前言部分的文献回顾与主要探索的四个目标之间，尤其是研究问题 4，在阐述过程中的因果衔接薄弱。

修改说明 5：感谢专家的宝贵意见，我们遵循您的建议，对前言相关部分进行了修改，以突出文献回顾与研究问题之间的关系。针对这个问题的修改部分较为零碎，我们不在这里一一赘述，烦请专家移步正文前言参考标橙文字。

意见 6：此外，（1）“自检报告”问题 8 提到了实验 1 剔除了 12 名被试数据及其原因，并在正文 2.3 部分分析了包括和不包括这些剔除被试的两种情况下的统计分析结果。然而，在正文 2.3 部分并没有相关的结果。（2）阐述清楚实验 1、2 和 3 给被试支付报酬的形式和数量。（3）de Kwaadsteniet et al. 2017, 2019 在参考文献列表中，应以姓氏的主要部分，即 Kwaadsteniet 进行排序。

修改说明 6：感谢专家的问题，在文中我们有些地方可能表述不够清楚，因此造成了一些歧义，我们在这里一一回答您的问题。1）在自检报告中，我们确实提到了在实验 1 我们剔除了 12 名被试的数据，因为他们未能通过有关实验说明的测试题，而在第一稿中，我们也对包括和不包括这 12 名被试的样本分别进行了统计分析，并显示结果无本质差异。但在修改稿中，由于专家 2 对实验操作提出了一些问题，为了使本文结果更有说服力，我们根据专家 2 的意见在 12 月重新设计了实验和收集了数据，在新的实验中，所有被试都通过了测试题，因此，在新的修改稿中并不存在剔除被试的问题。2）我们已按照您的建议，补充说明了实验报酬，并用橙色文字标出（分别位于正文 2.2 第 2 段、3.2.1 第 1 段和 4.2 第 2 段）。3）我们已遵循您的意见，重新排列了文献，并用橙色文字标出。再次衷心感谢您的建议，这对提升本文写作质量大有助益！

参考文献

陈思静, 何铨, 马剑虹. (2015). 第三方惩罚对合作行为的影响: 基于社会规范激活的解释. *心理学报, 47*(3), 389–405.

Bicchieri, C., Dimant, E., & Xiao, E. T. (2018). *Deviant or wrong? The effects of norm information on the efficacy of punishment* (PPE Working Papers 0016). Philadelphia, PA: Philosophy, Politics and Economics of University of Pennsylvania.

Cialdini, B., Kallgren, A., & Reno, R. (1991). A focus theory of normative conduct. *Advances in Experimental Social Psychology*, *24*, 201-234.

Cialdini, B., & Trost, M. (1998). Social influence: Social norms, conformity, and compliance. In T. Gilbert, T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, 151-192.). Boston, MA: McGraw-Hill

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology, 51*(3), 629-636.

Fehr, E., & Williams, T. (2018). *Social norms, endogenous sorting and the culture of cooperation*. (ECON Working Papers 267). Zurich, Switzerland: Department of Economics of University of Zurich.

Forquesato, P. (2016). Social norms of work ethic and incentives in organizations. *Journal of Economic Behavior & Organization*, *128*, 231-250.

Lois, G., & Wessa, M. (2019). Creating sanctioning norms in the lab: The influence of descriptive norms in third-party punishment. *Social Influence*, *14*(2), 50-63.

McDonald, R. I., & Crandall, C. S. (2015). Social norms and social influence. *Current Opinion in Behavioral Sciences*, *3*, 147-151.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York, NY: Springer

……

专家 2 的审稿意见及修改说明

作者针对上一稿意见逐一进行了详细认真的修改，重新收取了实验 1 的数据。经过本次修改，文章的整体逻辑和研究思路都更加清晰，实验设计也更加完善，同意该文章发表。但是几个小点需要修改。

意见 1：摘要应先简要介绍文章探讨的问题及研究目的，再讲如何解答这个问题，也即每个实验的内容。现在摘要收入有点突兀，建议补充。

修改说明 1：首先感谢专家对本文的肯定，这给了我们莫大的鼓舞！我们遵循两位专家的建议，在这一稿中重写了摘要。重写后的摘要如下：

第三方惩罚对合作的维系可能来自其经济功能，即惩罚降低了违规收益；但也有学者认为其规范提示功能同样重要——通过唤起个体的社会规范来提升合作。然而，由于先前研究没有区分惩罚的这两种功能，因而未能回答：当惩罚不足以影响违规收益时，是否还能有效促进合作？实验一（N = 252）操作了违规的不同收益，高收益组被试选择违规的预期收益总是大于等于无第三方的对照组，但高收益组的合作水平却显著高于对照组，这说明即使第三方惩罚无法降低违规收益，依然能抑制自利行为。实验二（N = 179）中相较于在第一阶段有第三方的独裁者博弈中未受惩罚的违规者，受罚的违规者在第二阶段无第三方的独裁者博弈中分配给对方的代币和公共物品博弈中投入公共账户的代币均表现出了更高的水平。中介模型检验显示，惩罚通过提升描述性和命令性规范有效促进了被试的合作水平。2（是否旁观惩罚）×2（旁观前后）设计的实验三（N = 179）显示，旁观惩罚后被试的合作水平显著高于旁观前，也高于未旁观惩罚的被试。与实验二类似，社会规范在惩罚与合作间也有显著的中介作用。这进一步证实惩罚对合作的促进在很大程度上是通过规范激活来实现的，并且存在两种溢出效应：按照社会规范聚焦理论，惩罚使合作的社会规范成为了被试当前的“焦点”，并抑制了曾经的违规者（纵向溢出效应）和旁观者（横向溢出效应）在新博弈情境下的自私行为，尽管在新情境下并不存在惩罚机制。这两种溢出效应的发现补充了文献中占主导地位的经济学解释，并为理解人类社会长时间、大规模的合作提供了新视角。

意见 2：格式：结果部分存在部分格式错误，请参照心理学报的要求修改。

修改说明 2：感谢专家的建议，我们已仔细检查全文，并改正了若干呈现格式上的错误。

第四轮

专家 1 的审稿意见：

经作者本次修改，文章对于摘要和理论铺垫介绍方面的表述更为清晰，建议发表。

回复：感谢专家对我们的肯定！几位专家的意见让本文的质量有了显著的提升，再次感谢！

编委复审意见：

意见：文章经过多次修改后效果很好，基本达到发表的水平。但是，在修改的过程中，审稿人和作者可能都忽视了学报对文章摘要的要求，现在的中英文摘要都过长，尤其是中文的，超出太多，必须修改。文章的研究结果介绍也可以适度删减，更突出重点。建议作者修改中英文摘要。

修改说明：非常感谢编委专家对文章整体质量的把控！我们已经按照您的意见认真修改了中英文摘要，在文中用紫色标记。

主编终审意见：

文章经过四轮修改质量有明显提高。该研究的实验设计很清晰明了，结果比较可靠。建议发表。

回复：非常感谢主编对本文的认可，这对我们是极大的鼓舞！