

《心理学报》审稿意见与作者回应

题目：第三方惩罚对合作行为的影响：基于社会规范激活的解释

作者：陈思静 何铨 马剑虹

第一轮

审稿人 1 意见：该研究通过两个实验研究力图探讨第三方惩罚对合作行为的影响。该课题立意新颖，实验构思巧妙，具有一定的理论意义和实用价值。现仅就方法和行文上的几个问题与作者商榷。

意见 1：在“2.1.3 实验变量与假设”的第一段，作者指出“实验一的因变量是个体的合作行为，操作定义是个体在独裁者博弈中分给接受者的钱数，个体分给接受者的钱越多，说明个体的合作行为水平越高”。请提供相关研究的证据。即，先前哪些研究用这种方法来测量合作行为。如果没有先前研究的证据，请说明为什么“个体在独裁者博弈中分给接受者的钱数”测量的是合作行为，而不是对他人利益的关心。

回应：①文中“合作行为”是一个抽象层面的概念，与“竞争行为”相对，独裁者博弈也是一个将现实抽象的博弈游戏，因此笔者认为独裁者分给接受者的钱数越多，其分配行为也就相对地越接近“合作行为”而非“竞争行为”。这是一个行为层面的分析，而非动机层面。②先前研究亦有类似做法，如 Daniel Nettle 等的文章中指出“Human cooperative behaviour, as assayed by decisions in experimental economic dilemmas such as the Dictator Game, is variable across human populations.” (Nettle D, Collóny A, Cockerill M. Variation in cooperative behaviour within a single city[J]. PloS one, 2011, 6(10): e26922.) ③基于上述两点，笔者在文中相应处进行了补充说明，并增加了引文。

意见 2：在“2.1.3 实验变量与假设”的最后，作者没有假设“描述性规范”和“命令性规范”有交互作用。请说明在“2.3.3 描述性规范和命令性规范对合作行为的方差分析”为什么包含交互作用一项。由于交互作用的存在，建议作者在引言部分和“2.1.3 实验变量与假设”的开始阐明理论上为什么描述性规范和命令性规范存在交互作用。

回应：根据评审建议在假设中增加了相应说明：由于命令性规范的识别需消耗更多的自我控制资源，描述性规范更易获得，特别是当命令性规范高而描述性规范低时，他人与命令性规范不一致的行为可能会让个体更倾向于遵循描述性规范。描述性规范不同时命令性规范对个体的影响可能会变化。并增添了存在交互作用的假设。

意见 3：在“2.3.3 描述性规范和命令性规范对合作行为的方差分析”中，由于交互作用项的存在，建议进行 simple slope test 分析，然后应着重讨论交互作用项。

回应：修改说明：由于两种规范均为分类变量（高/低），因此进一步对两种规范对合作行为的影响进行了简单效应分析，结果显示：描述性规范低时命令性规范对合作行为的影响不显著 ($F(1) = 0.46, p = .83$)，描述性规范高时命令性规范对合作行为的影响显著 ($F(1) = 11.87, p = .00$)；命令性规范低时描述性规范对合作行为的影响不显著 ($F(1) = 2.27, p = .13$)，描述性规范高时命令性规范对合作行为的影响显著 ($F(1) = 22.44, p = .00$)。对这一现象的解释是：个体出于自身利益最大化的考虑，可能会优先遵守较低的规范，因此当有一种规范低时，无论是何种规范，个体都会倾向于遵循该规范，而忽视另一种较高的规范。该部分内容已经

补充至文中相应处。

意见 4: 在“3.3 实验结果”开始，请报告控制检验（manipulation check）结果。

回应：修改说明：在文中相应处补充了控制检验的结果。

意见 5: 在“3.3.3 有第三方时社会规范对合作行为的影响”中的表 8，为什么分析中没有“命令性规范知觉”？

回应：表中“描述性规范知觉”应改为“命令性规范知觉”，此处系笔误。采用逐步回归分析两种规范对合作行为的影响，因为结果描述性规范没有进入模型，其对合作行为的影响不显著，所以在表中没有呈现。已经在文中对此进行了改正。

意见 6: 在“3.3.3 有第三方时社会规范对合作行为的影响”中的表 9 是第二轮结果，还是第三轮结果（表 9 的标题上是第三轮，表 9 下的一段说是第二轮）？

回应：此处系排版错误，应将“第二轮的逐步回归分析中，只有命令性规范知觉一个变量进入模型。回归分析显示，第二轮中命令性规范知觉对被试的合作行为产生了显著的影响，命令性规范知觉越高，被试的合作行为越高。”这一段提到表 9 之前，表 8 之后。已经在文中进行了改正。

意见 7: 在“3.3.3 有第三方时社会规范对合作行为的影响”的表 9 中，作者需要说明为什么在这个模型中描述性规范对合作行为有负面影响。

回应：描述性规范知觉对合作行为有负面作用，这可能有两方面原因：①在有第三方而且被告知第三方惩罚频率的情况下，个体的注意力更多放在如何避免被惩罚，描述性规范低的个体认为大多数人合作行为较低，这些个体在无第三方时的合作行为可能本身就较低，而在有第三方时由于更害怕被惩罚，反而会提高自己的合作行为。②第三方惩罚频率越高，个体知觉到的描述性规范也就越低，但由于较高的惩罚频率，个体自身受惩罚的预期概率也会提高，此时个体会更倾向于提高合作行为。以上两点已经在文中相应处进行了说明。

意见 8: 在引言部分，“1.1 社会规范与社会规范激活理论”之前，对研究的阐述不是很明确。

建议进行如下的调整：

A.研究的问题是什么？

B.为什么这个问题是重要的？

C.就这个问题已经有哪些研究？哪些点还没有人作研究？

D.本研究拟采用什么方法针对其中的一点进行研究。

回应：已经按照评审要求对该部分进行了修改。详见引言部分。

意见 9: 在“1.2 描述性规范、命令性规范与第三方惩罚”的第一段，第一次出现“描述性规范”和“命令性规范”时，最好能给出一个例子说明。

回应：已经在文中相应处进行了补充。

意见 10: 在“4.研究不足与结论”之前，作者最好总结一下本研究的发现和前人的研究有哪些一致，又有哪些不一致。和引言部分首尾呼应。

回应：已经将第四部分改为“讨论与结论”，补充说明了一致与不一致的结果及本研究的理论创新之处。

审稿人 2 意见：该论文通过两个独裁者博弈实验，考察第三方惩罚对合作行为的影响。研究发现合作行为在有第三方时主要受命令性规范影响；在无第三方时，主要受描述性规范影响。第三方惩罚频率越高，个体在有第三方时的合作行为也越高，撤去第三方后个体合作行为的下降也越大。该论文目的明确，逻辑清晰，结构合理，写作规范。但存在的问题及审稿人建议如下：

意见 1：1. 文献部分没有介绍以往研究对“描述性规范”和“命令性规范”这两个关键变量是如何测量和操作的。没有综合呈现“社会规范”的研究方向及其重要发现，理论阐述不够充分和深入，没有体现该研究有何创新之处。

回应：①关于两种规范的测量：以往研究对两种规范的测量通常是通过要求被试预期他人态度的方法来测量，并且针对不同的行为与情境，测量的题目也不同。如 Buunk (Buunk, B. P., & Bakker, A. B. (1995). Extradynamic sex: The role of descriptive and injunctive norms. *Journal of Sex Research*, 32(4), 313-318.) 等对滥交行为的研究中测量描述性规范的方法是：Respondents were asked to indicate the relative proportion of their friends and acquaintances who had extradyadic sexual relations. A seven-point scale was used with anchor points "no one," "some," "most of them," and "all of them."; 测量命令性规范的方法是：Respondents were asked to indicate on a nine-point scale ranging from "very objectionable" to "very good" what their best friends would think if they would engage in extradyadic sexual behavior. 这样的方法也常见于酒精成瘾、药物滥用、暴饮暴食、乱丢垃圾等方面的研究。通常描述性规范测量被试对周围人特定行为的预期，命令性规范测量被试对周围人针对特定行为态度的预期。在本研究中，由于被试是与陌生人进行博弈，因此用对大多数人行为的预期来测量描述性规范；由于针对的是第三方惩罚情境，因此用大多数人会对何种行为进行惩罚（大多数人反对何种行为）来测量命令性规范。在本文方法部分增加了上述对规范测量的说明。②关于社会规范研究方向及重要发现的综合呈现问题：在文献综述部分增加了对现有关于社会规范研究成果的评述。③关于本研究的创新之处：本研究的创新主要在于以下三方面。a) 以往关于两种社会规范的研究往往局限于某种特定行为，如乱丢垃圾、酒精成瘾等，本研究将其应用于合作行为这种更具有广泛意义和普遍价值的领域；b) 在对第三方惩罚影响合作的分析上，以往研究只看到第三方惩罚对合作的正面促进作用，本研究分析了第三方惩罚对合作的双重作用，不仅有促进合作的一面，也有抑制合作的一面；c) 以往对第三方惩罚的分析虽然也引入了社会规范，但没有涉及两种不同的社会规范，本研究分析了第三方惩罚影响合作的双重作用的心理机制，用两种规范的不同作用来解释这样的双重作用。已经在“讨论与结论”部分补充说明了本研究的理论创新点。

意见 2：方法部分，实验 2 设置了 0~90% 第三方惩罚的频率，该设置的高频率与社会现实差距较大，缺乏真实性，如果被试对此操作产生怀疑，就会影响到实验结果，有必要对此进行操作检验。

回应：①第三方惩罚频率 0~90% 的设置出发点是为了在实验中穷尽可能的范围，以便收集更全面的数据进行分析比较。的确存在生态效度问题，但笔者认为作为一个实验研究，在数据的连贯性上还是具有一定的价值。②关于第三方惩罚实际发生的频率，有研究指出，合作者与竞争者同时存在时，对竞争者的第三方惩罚频率可高达 45.8%，并且接受者对第三方惩罚的预期还高于此频率 (Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and human behavior*, 25(2), 63-87.)，我国学者范良聪等的研究发现第三方惩罚发生的频率为 55.4% (范良聪, 刘璐, 梁捷, (2013). 第三方的惩罚需求：一个实验研究, 经

济研究, 5: 98-111.); 谢娉发现对于“不公平且低于规则”的分配, 第三方惩罚发生的频率高达 76.9% (谢娉, (2013). 第二方惩罚与第三方惩罚的产生机制差异比较--基于情绪和规则的研究. 浙江大学硕士论文.). 这说明在某些情况下较高的第三方惩罚也是有可能发生的。③即使只考虑 0-50% 频率的第三方惩罚, 依然可以得出本研究的主要结论。出于上述原因, 在文中对此问题进行了说明, 并在实验结果中增加了操作检验。

意见 3: 实验二的实验程序介绍太简单, 信息不够全面, 建议按实验步骤来介绍, 包括什么时候收集知觉到的“描述性规范”和“命令性规范”, 共收集多少次。

回应: 已经在相应处补充介绍了实验二的实验程序。

意见 4: 结果部分, 本文的两个实验分别收集了 120 和 300 个大学生被试, 但在统计分析中没有考察专业和性别差异, 作者有必要交代这些变量对结果是否存在影响。建议作者进行慎重分析和解释。

回应: 在本研究中并未发现性别和专业对合作行为存在显著影响, 并且以往研究中也少有此类发现, 因此未在文中呈现相关统计分析结果。笔者以性别和专业为自变量、以合作行为与两种规范为因变量分别做方差分析, 结果均不显著, 详见下表:

表 0-1 实验一性别和专业对各因变量的方差分析

因变量	<i>df</i>	<i>F</i>	<i>p</i>
第一轮合作行为	2	.28	.76
第二轮合作行为	2	.25	.78
第一轮描述性规范	2	.03	.97
第二轮描述性规范	2	1.05	.35
第一轮命令性规范	2	2.91	.06
第二轮命令性规范	2	1.47	.24

表 0-2 实验二性别和专业对各因变量的方差分析

因变量	<i>df</i>	<i>F</i>	<i>p</i>
第一轮合作行为	2	.27	.76
第二轮合作行为	2	.95	.39
第三轮合作行为	2	.70	.50
第四轮合作行为	2	.51	.60
第一轮描述性规范	2	.73	.48
第二轮描述性规范	2	.24	.78
第三轮描述性规范	2	.96	.38
第四轮描述性规范	2	1.46	.23
第一轮命令性规范	2	1.15	.32

第二轮命令性规范	2	.95	.39
第三轮命令性规范	2	.96	.39
第四轮命令性规范	2	.27	.76

据此在两个实验的结果部分增加了对性别、专业变量未影响结果的说明。

意见 5: 第三方惩罚对合作的影响过程可能存在很多因素, 比如社会责任感, 公平感以及情绪的影响。该研究是否排除了这些变量的影响? 建议作者适当补充一些说明和解释。

回应: 本研究用社会规范来说明第三方惩罚影响合作的心理机制, 主要原因是社会规范两种不同形式(描述性、命令性)为第三方惩罚对合作的促进与抑制双重作用提供了较好的解释。而社会责任感、公平感等变量虽然能够解释第三方惩罚对合作的促进作用, 但较难说明为什么惩罚撤除后合作行为会下降到先前没有惩罚的基线以下。情绪在第三方惩罚影响合作的过程中应该发挥了重要作用, 但目前研究经常涉及到的特定情绪, 如愤怒对第三方惩罚的影响、内疚与后悔对合作行为的影响都是只涉及一种行为, 并没有解释第三方惩罚影响合作的整个过程。当然, 该过程中可能涉及不止一种、或是复合型的社会情绪, 这确实是亟需研究的一个重要问题, 但限于篇幅和技术手段, 本文很难做出进一步分析, 因此将以上几个变量的分析作为研究局限和未来研究的一个建议在讨论部分进行了说明。

意见 6: 总讨论部分过于简略, 作者只关注研究结果的讨论, 没有说明该研究方法和结果是否能够鉴别和解释前人研究的结果及其关系, 建议联系同类研究结果进行比较, 适当讨论该研究的理论贡献和运用价值。

回应: 已经将第四部分改为“讨论与结论”, 讨论了与前人研究一致和不一致的情况及可能的解释。并增加了理论贡献和运用价值的分析。详见相应处翻红文字。

意见 7: 论文中多处存在字词错误, 对理解造成混乱。错误出现处包括 2.1.3; 3.1.2; 3.2.2; 3.3.1。

回应: 抱歉, 已经核对并修改。

意见 8: 参考文献中, 有部分文献重复, 如第 2 和第 3 个文献是同样的。建议仔细修改。

回应: 已将重复处删除, 并核对了参考文献部分。

第二轮

审稿人 1 意见:

意见 1: 在“1.引言”的第一段, 作者指出“在实验室情景中也是一种可以跨越多种博弈情境的现象。第三方博弈可以与独裁者博弈(Fehr & Fischbacher, 2004a)、信任博弈(Ferrin, Dirks & Shah, 2006)、社会两难博弈(Fehr & Fischbacher, 2004a)等不同的博弈情境相结合。本研究旨在借助上述博弈情境对第三方惩罚影响合作行为的心理机制进行分析, 并对社会规范在此过程中的作用进行探讨。”这几句话应该是作者回应审稿人的建议。可是, 放在此处打断了原有的逻辑。建议删除。

回应: 已经根据您的意见将该段文字删除。

意见 2: 在“1.1 社会规范与社会规范激活理论”上面的两段, 作者指出: “这些研究虽然指出了社会规范在第三方惩罚中的重要作用, 但并没有对不同社会规范的影响作出进一步比较分

析。”因为该处是第一次出现“不同社会规范”这个术语,最好简单地说明是哪些不同种类的社会规范。而且,此处还应说明为什么“对不同社会规范的影响作进一步比较分析”是重要的。回应:已经根据您的意见在此处增加了以下文字:“社会规范除了代表“应如何做”的命令性规范(injunctive norms),还有代表“大多数人如何做”的描述性规范(descriptive norms)。有关第三方惩罚的多数研究所涉及的社会规范都是命令性规范,但社会心理学对两种规范的研究表明,很多情况下描述性规范对个体的影响更大(Cialdini, 1991)。”

意见 3: 在“1.3 第三方惩罚与合作行为”一节,作者认为先前有关第三方惩罚与合作行为的研究是用自我决定理论来结实其内部机制的。作者在这一节中最好再深入一步说明自我决定理论和社会规范理论的联系。这样可以使逻辑思路更加流畅,没有脱节之感。

回应:感谢您的意见,这对深化研究内涵、引出理论假设和贯通文章思路都大有帮助。已经根据您的意见在此处最后增加了一段说明文字:“自我决定理论为惩罚对合作的双面作用提供了一个动机层面的解释,但第三方惩罚作为一种纯粹的利他行为,与上述很多研究中作为一种机制的惩罚(sanction)有很大的不同。第三方惩罚除了作为一种外在动机推动个体的合作行为,还可能对个体社会规范的唤起和激活,在这个层面上,也许社会规范激活理论对第三方惩罚影响合作行为的解释更贴近其利他行为的本质。如果撤除第三方惩罚之后个体的合作行为显著下降,那么说明自我决定理论的解释是充分的;如果合作行为没有出现下降,那么说明第三方惩罚激发了个体对社会规范的感知,这更符合社会规范激活理论的解释。”

意见 4: 本文的标题是“第三方惩罚对合作行为的影响:基于社会规范激活的解释”。读者会期望该研究的自变量是第三方惩罚,因变量是合作行为。内在机制是社会规范激活。可是,实验一的自变量是两种不同的社会规范,因变量是合作行为。实验二的自变量是第三方惩罚,因变量是合作行为。作者在“2 实验一:描述性和命令性规范对合作行为的影响”的前一段应详细说明两个实验之间的内在逻辑。即,实验二是如何建立在实验一的基础上探索内部机制的。

回应:两个实验之间逻辑关联方面表述的缺失的确妨碍了对整个研究思路的理解。已经根据您的意见在两个实验之前增加了一段说明其逻辑关联的文字:“研究试图通过分析两种社会规范对个体的不同影响,在一定程度上揭示第三方惩罚影响合作的心理机制。因此实验一对两种社会规范影响个体合作行为的分析是实验二研究第三方惩罚影响合作行为的基础。如果两种社会规范对合作行为的影响的确不同,而这种不同又与第三方惩罚对合作的双面影响产生一致和对应的关系,那么就有理由相信第三方惩罚影响合作行为的心理机制与两种社会规范的不同作用存在很大关系。”

意见 5: 在“2.3.5 描述性规范知觉与命令性规范知觉对合作行为的回归分析”中,为什么没有包括交互作用项?而且,在该回归分析中,描述性规范知觉与命令性规范知觉都是二值变量,为什么可以通过标准偏回归系数来比较相对影响?

回应:这是本研究的一个方法上的疏忽,为此在该处使用带有交互项的多元回归分析重新进行了统计,并修改了原有表述。

意见 6: 在“2.4 讨论”的第二段,作者指出:“首先被试由于害怕被第三方惩罚,在做出较低分配时会产生害怕的情绪;其次被试由于命令性规范的激活,在做出较低分配时,会因为自己的行为违背了公平规范而产生内疚的情绪”。这是一个很好的推测。如果有实验三直接检验这个假设,那么就更有说服力了。

回应:感谢您的意见,这为进一步的研究指出了很好的方向。本研究主要是着眼于第三方惩

罚对合作产生影响的心理机制，因此没有对这个解释进行进一步的分析和验证。事实上，这种推测也是建立在现有研究的基础上，第三方惩罚是一个社会规范激活的过程，这已有研究提供了相关证据（陈思静等，2011）；而个体违背社会规范时会产生内疚情绪，也已为相关研究所证明（Azar，2007）。针对这一问题，我们已经在文中相应表述处增加与之相关的参考文献说明，以增强其说服力。

意见 7：实验 2 的第一轮，作者应明确说明结果是否得到了和实验一一致的模式。另外，由于实验一中已经发现描述性规范知觉与命令性规范知觉存在交互作用，为什么在实验二的回归分析中没有包含交互作用项？

回应：在对实验二第一轮的分析中确实应使用带有交互项的多元回归统计，已经在该处进行了修改并重新报告了结果，同时明确了该结果与实验一具有一致的模式。

审稿人 2 意见：

意见 1：1.2 部分的小标题是：描述性规范、命令性规范与第三方惩罚。但这一节并没有提及第三方惩罚，需要调整。

回应：感谢您的细心阅读，已经根据您的意见将该处标题改为“1.2 描述性规范与命令性规范”。

意见 2：2.3.1 和 3.3.1 部分检验了被试的性别和专业对实验结果的影响，文中可以只报告这两个变量对结果并没有显著影响，而表 1 和表 7 的分析结果可以不报告。这样可以节省篇幅。

回应：已经根据您的意见将该处表格删去，并且在报告变量对结果没有显著影响时注明了 F 值和 p 值的分布范围。

意见 3：修改稿对有些分析结果图表并用，占用很大篇幅。例如，表 14 和图 2，表 15 和图 3。建议对一个分析结果只保留图或表来说明。

回应：已经根据您的意见将该图 2 和图 3 删去，仅保留 SNK 检验的表格。

意见 4：4.2 部分的结论过长，和讨论无异。内容与前面的讨论基本重复，建议将此部分简略归纳为几句话比较符合结论的格式。

回应：已经根据您的意见将结论部分进行了修改。

意见 5：3.1.4 部分第 10 行，“……具有较高的合惩罚撤除。这句话词句不通。建议修正。

回应：已经根据您的意见将该处语句进行了修改，修改后的语句是：“所以虽然个体此时惧怕惩罚，而遵从命令性规范，具有较高的合作水平；但一旦惩罚撤除，命令性规范强度降低，个体的合作水平会迅速下降。”。

审稿人 3 意见：

意见 1：本研究的新贡献是什么？这一点作者需要更为清楚的加以说明。关于描述性规范和命令性规范的区别，社会心理学中 Cialdini 等人已提出并加以考察，作者也引用了。那么关于这两种规范的研究中，究竟有什么未知或有争议的问题，是本研究试图解决的呢？关于第三方惩罚，行为经济学中也有一定数量的研究。惩罚不一定总是带来正面的结果，也有可能带来负面的结果，这一点也有不少的研究指出。这其中究竟有什么未解决的问题，本研究对

该问题的新的贡献究竟在哪里，需要作者说明。

回应：感谢您的质疑，这敦促我们进一步思考本研究的价值。我们认为本研究的新贡献主要体现在以下几个方面：①两种社会规范的区别的确已有不少研究，并已产生一些有影响力的理论。现有研究大多数着眼于两种规范分别作用于个体时的不同结果、在激活特定类型规范时个体的不同反应，但在同一实验中分析两种社会规范冲突时（本研究中为一高一低）个体合作行为模式的研究还很少。本研究实验一的一部分结论与以往对两种社会规范的研究相符，如描述性规范对个体的影响更大。但还有一部分结论是不同于以往研究，或者说发展了现有研究结论的。实验一发现当两种社会规范不同时，个体都会倾向于遵守成本较低的规范，而与规范类型无关。由此我们可以假设个体在面对不同社会规范时可能遵循“下限法则”，即优先遵循成本较低的规范。在符合了最“低”社会规范的前提下，个体更倾向于按利益最大化原则行事。这一推论对于两种社会规范对合作行为的影响有了新的补充。实验一可以说一定程度上回答了：两种社会规范不同时个体会怎么做？而目前对于该问题，Cialdini 等提出的社会规范聚焦理论的回答是：哪种规范激活程度高个体就遵循该规范。本研究的结果并未支持这一理论，而是倾向于用上述“下限法则”来解释。当然，这还需要进一步研究的补充和验证。由于实验一的目的是为第三方惩罚影响合作提供心理机制解释的基础，所以在文中对上述结果未做进一步分析，但我们认为这一结果为后续研究提供了一个较有新意和价值的思路。②第三方惩罚的研究主要有三类：一是研究第三方惩罚现象的存在及其意义；二是研究第三方惩罚发生的原因、个体差异、影响其发生的内外部因素等，可以说是以第三方惩罚为因变量的研究；三是第三方惩罚对其他因素的影响，可以说是以第三方惩罚为自变量的研究。目前第三类研究较少，而且这其中的多数研究只是表明第三方惩罚对社会规范或合作有促进作用或积极影响，并未分析影响的具体过程。本研究不仅关注第三方惩罚对合作行为的影响，并且试图进一步分析这种影响的心理机制，这在第三方惩罚研究，尤其是行为经济学的研究中并不多见。本文试图在一定程度上回答：为什么第三方惩罚能够促进合作行为？这一问题目前并未有任何公认的答案。③惩罚并不总带来正面结果，这一现象之前的研究早有涉及。但第三方惩罚作为一种纯粹的利他行为，分析其对合作负面影响的研究并不多见。我们认为第三方惩罚虽然也属于惩罚，但由于其特殊性，对其影响合作的双面作用的分析虽不能说有创新，但也不能说无价值。更重要的是，目前对惩罚双面作用的解释主要是基于自我决定理论，即内部动机被外部动机所排挤。但本研究实验二发现，10%的第三方惩罚频率撤除后个体的合作行为并未下降，这不符合自我决定理论的预期。因此我们用社会规范激活理论来解释这一现象，进而用两种规范不同条件下的激活来解释第三方惩罚对合作行为的双面影响，这在一定程度上是对已有理论的发展，而且实验二的发现，不管在第三方惩罚领域还是惩罚机制领域，都是一个新颖和有价值的现象。为了更清楚说明这一点，我们在论文1.3 第三方惩罚对合作行为的影响这一部分增加了一段话：“自我决定理论为惩罚对合作的双面作用提供了一个动机层面的解释，但第三方惩罚作为一种纯粹的利他行为，与上述很多研究中作为一种机制的惩罚（sanction）有很大的不同。第三方惩罚除了作为一种外在动机推动个体的合作行为，还可能对个体社会规范的唤起和激活，在这个层面上，也许社会规范激活理论对第三方惩罚作用于合作行为的解释更贴近其利他行为的本质。如果撤除第三方惩罚之后个体的合作行为显著下降，那么说明自我决定理论的解释是充分的；如果合作行为没有出现下降，那么说明第三方惩罚激发了个体对社会规范的感知，这更符合社会规范激活理论的解释。”④诚如评审一开始就指出的，本研究试图把两种社会规范与第三方惩罚的研究联系起来，如果仅仅是方法和形式上的叠加，那么不能说有什么创新和价值，但本研究用两种社会规范的不同作用来解释第三方惩罚对合作的双面影响，这样的对应关系产生了形式和内容两方面的结合，为以后的研究提供了一个较有新意的视角。

以上是我们在对本研究新贡献的几点理解，或许存在一定程度的“实验者效应”，但基本都是

基于对现有问题的思考，恳请评审专家予以考虑。

意见 2：在“描述性规范”和“命令性规范”这一对范畴上，存在着一些概念上的问题，需要思考和澄清。最核心的问题是，一种行为模式，如果群体中的大多数人都不遵守，它还是一种群体规范吗？还会有群体内部对于它的第三方惩罚吗？例如，如果一个群体或者社会中的多数人，在日常行为中都是送礼行贿的，那么还能说这个群体或社会的规范是不行贿吗？尽管可能有法律禁止送礼行贿，但是这恰恰说明群体的非正式规范和外在的正式法律是背离的。当然，更困难的问题是如果群体中的大部分人都认为送礼行贿从道德角度看是不太好的，但是在行为上大家都送礼行贿，那么这个群体中关于送礼行贿的“规范”到底是什么？更进一步，如果在行为上大家都不遵守该行为模式，那么（无论是否仍然将该行为模式称作命令性规范）大家还会自发的展开针对违反者的第三方惩罚吗？尤其要注意的是，规范和法律不同，其所依赖的第三方惩罚不是由群体之外的执法者强加的，而是群体内部参与博弈者之间自发开展的。如果 90%的博弈参与者自己都不按照该模式行动，群体内部的第三方惩罚在现实中有多少可能性会发生？

回应：您提出的是一个很有价值且耐人寻味的问题，下面我们根据自己的理解试着逐一回答您的问题。①“一种行为模式，如果群体中的大多数人都不遵守，它还是一种群体规范吗？还会有群体内部对于它的第三方惩罚吗？”诚如您所质疑的，从概念内涵来讲，群体规范本应是指群体内大家都倾向于遵守或执行的行为准则。但不论从理论层面还是现实角度，群体“推崇”或“标榜”的规范与实际上群体所“奉行”或“热衷”的规范之间经常有很大的差别。事实上，描述性与命令性规范这一区别的提出，本就是为了解决这一问题（Cialdini, Kallgren & Reno, 1991）。我们认为，首先，群体中“应当如何做”和“大多数人如何做”是两个不同层面的规范，前者是一个价值判断，后者是一个事实判断。当然两者可能相互影响，但绝对是不能等同的，如果两者在绝大多数情况下都是一致的，也就没有区分它们的必要了。其次，不管描述性规范高还是低，命令性规范是相对独立而明确的。比如，可能红灯时过马路的人很多，但不能说因为这样“红灯停，绿灯行”的规范就改变了。命令性规范是群体中大多数人赞成或反对的，这是一种价值判断，也许当多数人不遵从该规范时，由于“法不责众”，该规范对人们的影响会减小。但不能说就此这一规范就被改变或撤销了。同样的，也许很多人认为在独裁者游戏中分给对方少一点也无伤大雅，但这并不能代表这一情境中的公平规范就改变了。最后，第三方惩罚其实是个体对不符合自己价值判断的行为进行惩罚，因此即使在独裁者游戏中所有人都倾向于少分给对方，这也不能说明这种行为就符合了第三方的价值判断，只能说这给了低分配行为一个很好的道德推脱方式——大多数人都是这样做的。但这并不能说这样的行为就是道德的，或者符合命令性规范的。如果一个个体有稳定的价值判断，就很可能对这样的行为做出惩罚。在现实社会中，很多不道德行为，比如您所举例的行贿受贿，都成了大家司空见惯的现象，但还是会有很多人站出来举报、上访，甚至有些是没有涉及自身利益的。这说明即使描述性规范与命令性规范相距甚远，个体还是可能因为违背命令性规范而遭到惩罚。②“例如，如果一个群体或者社会中的多数人，在日常行为中都是送礼行贿的，那么还能说这个群体或社会的规范是不行贿吗？”多数人会行贿受贿，说明社会中描述性规范是行贿受贿。但这不能说明大多数人赞成这一行为。我们认为不会有人认认为行贿受贿是道德的、被社会赞成的。因此命令性规范依然是不能行贿受贿。这是一个典型的描述性规范与命令性规范冲突的例子，这样的例子恰恰说明了实验一对两种规范不同时个体合作行为的研究是很有意义的。我们认为不能说大家都这样做，这种行为就是公认正确或是应该的了。事实判断与道德判断在这样的情境中是相对独立的。③“如果在行为上大家都不遵守该行为模式，那么（无论是否仍然将该行为模式称作命令性规范）大家还会自发的展开针对违反者的第三方惩罚吗？如果 90%的博弈参与者自己都不按照该模式行

动，群体内部的第三方惩罚在现实中有多少可能性会发生？”关于这一问题，在上面已有论述。另外，关于第三方惩罚实际发生的频率，有研究指出，合作者与竞争者同时存在时，对竞争者的第三方惩罚频率可高达45.8%，并且接受者对第三方惩罚的预期还高于此频率(Fehr, E., & Fischbacher, U. (2004). *Third-party punishment and social norms. Evolution and human behavior, 25(2), 63-87.*), 我国学者范良聪等的研究发现第三方惩罚发生的频率为55.4% (范良聪, 刘璐, 梁捷, (2013). 第三方的惩罚需求: 一个实验研究, *经济研究, 5: 98-111.*); 谢娉发现对于“不公平且低于规则”的分配, 第三方惩罚发生的频率高达76.9% (谢娉, (2013). 第二方惩罚与第三方惩罚的产生机制差异比较--基于情绪和规则的研究. 浙江大学硕士论文.)。这说明在某些情况下较高频率的第三方惩罚也是有可能发生的。

综上所述, 我们认为两种规范是可以相对独立存在并分别作用于个体行为的。在我们的社会中, “潜规则”无疑是普遍存在的, 但这并不能够替代显性的道德规范。康德曾说过: “有两种东西, 我对它们的思考越是深沉和持久, 它们在我心灵中唤起的惊奇和敬畏就会日新月异, 不断增长, 这就是我头上的星空和心中的道德定律。”虽然康德在私生活里并不是个“圣人”, 但这并未妨碍他在认知里对道德规范的推崇。也许我们永远做不了一个时时处处符合“命令性规范”要求的“圣人”, 甚至因为“描述性规范”而不断地刷新道德底线, 但这些并不能阻止我们对道德定律“虽不能至, 心向往之”的价值认同。

意见 3: 如果, 如上所述, 任何“命令式规范”都其实隐含了“描述性”的信息 (即大多数人是遵守该行为模式的), 那么将两种规范看作是可以独立加以测量和操作的概念就是不适当的。当研究者操作“命令性规范”时, 其实已经隐含地传达给被试描述性的信息。而当研究者告知被试描述性信息时 (例如 90%的人都遵守或者不遵守某一行为模式), 命令性规范也显然会受到相应的加强或者削弱。

回应: 我们认为, 命令性规范在一些情境中可能隐含了描述性信息, 但也有很大可能是与描述性规范冲突的。比如不在公共场合吸烟是一个命令性规范, 但这并不能说明大多数人会这样做。同样的, 在独裁者博弈中, 第三方惩罚频率的增加说明违背公平规范的人很多, 但这恰恰说明描述性规范低。因此这样的情境中描述性规范与命令性规范是负相关的。而在第一轮没有第三方的情况下, 如评审所说, 描述性规范与命令性规范是正相关的。这样的关系也说明这两种规范相互虽有关联, 但却是相对独立的。

意见 4: 在实验一中, 作者用触发第三方惩罚的阈值作为命令式规范的操作化和测量。这中间也有概念问题。命令式规范从概念上讲只是指明在一定情境中“适当的”或者“良好的”的行为模式是怎样的, 并不一定意味着不遵从该模式就一定会受到群体内的第三方惩罚。例如, 向慈善机构捐款是符合社会规范的, 但是不捐款也不会受到第三方惩罚。具体到独裁者博弈这一情境中, 分配给对方一定的资源固然可以看作是好的、亲社会的行为, 但是并没有清晰的、可参考的社会规范要求说对于怎样的“自私”行为就需要惩罚。

回应: 这是一个概念操作上的问题, 命令性规范严格地来讲, 是“群体中大多数人赞成或反对的行为” (Cialdini, Kallgren & Reno, 1991)。由于我们探讨的是第三方惩罚, 因此采用了“大多数人会反对何种行为”这样的操作和测量。这样测量的原因主要是: 不受到第三方惩罚, 也就意味着个体觉得社会中大多数人不会反对这样的行为。这可以说是一个行为的“底线”。因为在博弈环境中个人利益导向性比较强, 因此我们倾向于测量个体能够接受的符合规范的最低成本行为。而且这样的测量也相对而言降低了社会称许性的影响。

意见 5: 实验一 (以及实验二) 的程序部分需要写得更为具体、翔实。例如, 独裁者博弈的指示是如何措辞的。被试何时回答关于描述性规范认知和命令性规范认知的问题。费用如何

发放。等等。

回应：我们根据您的意见补充了原先对实验程序的描述。

意见 6：实验一结果中，应当首先报告第一轮的结果，包括被试的实际行为、对描述性规范的认知、对命令性规范的认知、行为与规范认知之间的相关性，作为和第二轮比较的基础。

回应：已经根据您的意见在该处补充了相关系数的说明。

意见 7：从表 5 推测，在第一轮中，低-低组的平均分享行为有 2.43，而低-高组的平均数为 1.83，高-低组 1.93，高-高组 1.49。如果 pre-test 有如此大的差异，简单看 post-test 的各组差异的意义就不大，更要看每一组的 pre-test 和 post-test 的前后差异。对此作者应该给予更清楚的说明。

回应：非常感谢您的细心阅读，我们根据您的意见补充了对四组第一轮和第二轮合作行为差异的检验。

意见 8：实验一表 6 分析第二轮中规范知觉对行为的影响，意义不明显。规范知觉对于行为的影响应当在第一轮中最重要的，第二轮中如果要分析，应当分析对规范的操作是否通过影响相应的规范知觉而影响行为。

回应：非常感谢您的评审意见，这是研究方法上的一个疏漏，已经将其修改，直接分析规范操作对行为的影响。

意见 9：实验二的程序部分描述非常不充分。尤其是关于惩罚。在惩罚发生的第二轮，被试被告知的惩罚的信息有哪些？谁是惩罚者？惩罚者是否也参与独裁者博弈？惩罚者按什么规则进行惩罚？惩罚有无上限？而且，究竟实际上有没有被试受到惩罚？如果有，哪些被试受到惩罚？如果没有，怎么能够使得被试(尤其是那些分给对方 0 元的)相信，惩罚是真实存在的？第三轮中，被试被告知的其他被试所受惩罚的信息，除了百分比之外，有没有其他信息，例如触发惩罚的阈值，惩罚的力度（损失大小）等？被试有没有收到关于上一轮的描述性规范（即平均分配行为）的信息？

回应：我们试着逐一回答您的问题。

在惩罚发生的第二轮，被试被告知的惩罚的信息有哪些？惩罚者按什么规则进行惩罚？惩罚有无上限？回答：关于这一点在实验范式中已经说明，在第二轮，主试告知被试本轮存在第三方，第三方有惩罚的权力，其初始钱数为 5 元。第三方每付出 1 元，可以使被惩罚者减少 2 元。

谁是惩罚者？惩罚者是否也参与独裁者博弈？回答：关于这一点在实验程序中已经说明，首先进行身份的分配。分配的角色有独裁者博弈中的分配者、接受者和第三方博弈中的第三种，在实验中呈现的角色名称为 A/B/C 三种。每位被试都被分配为 A 角色（分配者），然后是对每个角色的说明，被试被告知同组的另两名被试分别扮演 B（接受者）和 C（第三方）角色。

而且，究竟实际上有没有被试受到惩罚？如果有，哪些被试受到惩罚？如果没有，怎么能够使得被试(尤其是那些分给对方 0 元的)相信，惩罚是真实存在的？回答：被试不会收到惩罚结果的反馈信息。事实上，每轮分配或惩罚的结果都不会反馈，被试在事先就被告知这一点，其实验报酬是在四轮中随机抽取一轮的分配结果，在全部实验结束后反馈给被试。

第三轮中，被试被告知的其他被试所受惩罚的信息，除了百分比之外，有没有其他信息，例如触发惩罚的阈值，惩罚的力度（损失大小）等？被试有没有收到关于上一轮的描述性规范（即平均分配行为）的信息？回答：凡是被试收到的反馈信息，实验程序中都做了说明，没

有说明的即是没有反馈。

根据以上的意见对实验二程序说明进行了修改和补充。

意见 10：实验二 3.3.3 和 3.3.4 两部分，分析两种规范的知觉对于行为的影响，逐步回归是不适当的。两种规范知觉之间的相关性是否较强？应当使用多元回归，将两种规范知觉同时纳入回归方程。

回应：我们根据您的意见用多元回归重新分析了结果。

意见 11：表 11 的分析中，还应当放入被试收到的上一轮惩罚的信息作为自变量。

回应：对实验二第三轮的回归分析进行了修改，重新用多元回归分析。由于我们的理论假设是第三方惩罚频率通过影响描述性规范知觉来改变其行为，因此在 3.3.5 部分最后还补充分析描述性规范知觉在惩罚频率与合作行为之间的中介效应。

意见 12：第四轮的结果显示，当被试得知，有大量的其他参与者受到惩罚时，当惩罚被撤除之后，被试降低了自己的分享行为。我同意作者的解释说这是因为被试认识到，大多数人其实分给对方的金额是较低的。但是我认为作者将这一点看作是仅仅降低了“描述性规范”是不妥的。如前所述，当个体认识到大多数人都不遵守某一行为模式时，该行为模式能否还能成为“命令性规范”也成为问题。所以大多数人的“越轨”行为不仅仅削弱了该行为的描述规范性，也削弱了其命令规范性。甚至可以进一步说，当个体认识到某种行为模式并不被群体中的大多数人所遵循，但是却被某种外在力量（在实验中，第三方惩罚者并不是独裁者博弈的参与者）所强制维持，该行为模式的正当性基础会受到动摇（如果大家都和我一样，只分给对方 1 块钱，凭什么那些惩罚者要强迫我们分给对方 3 块钱？）。所以当外在的强制力量撤出后，由于该行为模式无论作为描述性规范还是命令性规范都失去了合法性基础，也就不再获得遵守。

回应：非常感谢您的评审意见，关于“当个体认识到大多数人都不遵守某一行为模式时，该行为模式能否还能成为“命令性规范”也成为问题。”这一问题已在前面根据我们的理解做出了回答。关于评审认为命令性规范包含描述性信息的问题，也已根据我们的认识在前面有所论述。在本研究中，我们认为描述性规范是事实判断，大多数人会怎么做；而命令性规范是价值判断，大多数人赞成或反对何种行为。在实验中被试接触到的只是一个小的群体，这样的行为概率并不足以推翻被试建立在以往经验上的价值判断。另外，从几轮合作行为与规范知觉的图上可以看到，惩罚撤除后的合作行为与描述性规范的趋势相一致，与命令性规范趋势不符。在惩罚撤除后，命令性规范并没有降低。

第三轮

编审意见：

意见 1：图 1 描述了描述性规范和命令性规范的交互作用。对理解这个交互作用很重要。表 5 中显著的描述性规范知觉*命令性规范知觉交互作用和表 7 中显著的描述性规范知觉*命令性规范知觉交互作用也需要用图描述，以帮助读者明白交互作用的形态。

回应：为了更直观显示两种规范的交互作用，的确很有必要呈现回归分析中的两个自变量交互作用图。我们根据编委专家的要求在文中相应处补充了回归分析的交互作用图。

意见 2：表 8 和表 9 的回归分析中，没有像表 5 的分析中同样加入描述性规范知觉和命令性规范知觉的交互作用。会否跟之前的分析不一致？

回应：根据专家意见我们在表 5 的回归分析中加入了交互项的检验，对于表 9 的回归分析，我们考虑到：1) 因为有三个自变量，其交互作用较复杂并难于解释；2) 表 9 的分析中我们主要希望验证在该轮次第三方惩罚频率对合作行为的影响比两种规范更大，根据回归分析中三个自变量的主效应已经足以支持这一结论；3) 关于两种规范的交互作用和第三方惩罚对合作行为的影响在之前和后面的分析中都分别有具体地展开。因此我们在该处没有增加交互项的分析。