

《心理学报》审稿意见与作者回应

题目：多元文化经历提升人类对机器人的利他行为及心智知觉的中介作用

作者：滕玥 张昊天 赵偲琪 彭凯平 胡晓檬

第一轮

审稿人 1:

本文从社会心理学的视角考察了人机关系,分析了多元文化经历对人类是否将机器人视为有心智的主体,并对其采取帮助行为的倾向。整体而言,该研究对了解影响人机关系的因素具有重要的理论意义。本人有若干建议和问题与作者团队探讨。

作者回复:十分感谢专家对拙作提出的非常详尽的建议!专家的若干建设性意见让我们深深认识到初稿存在的诸多不足之处。根据专家的意见,我们对初稿进行了大规模的修改:

在理论上,我们结合专家的意见重新整合梳理了多元文化经历促进对机器人利他行为的心理机制,并更为详细地论述了多元文化经历如何促进对机器人的心智知觉,进而影响对机器人利他行为的中介路径。

在实证上,我们重新补做了实验 2。优化了对机器人利他行为的测量方式,将图片呈现改为了自陈量表,并加入了积极与消极情绪作为协变量,排除了可能的替代解释。

我们对您的意见进行逐条回复如下,并在文中采用了修订模式修改全文,新增加的内容均已标红,您可以看见本文完整的修改过程。最后,再次感谢专家对拙作的批评指正,您的意见使得我们的稿件有了质的提升。

一、理论问题

1. 整体而言,作者对关键概念的选择、采取的理论视角、概念之间的逻辑关系解释的还不够清楚。

(1) 关于道德考量与利他行为

作者在引言第二段引入了“道德考量”,提出机器人可以被视为具有“道德地位的实体”,但并没有界定或解释清楚什么叫道德考量,什么叫有道德地位,实体是什么。为什么拥有道德地位的实体,人们就会对其做出利他行为?另外,第二段里的道德似乎跟第一段中提到的道德伦理问题中的道德不是一个范畴。上一段介绍的更多是使用技术与机器如何影响(甚至威胁)人类、使用智能技术中的伦理问题(比如是否可以给人脑外接智能芯片、人类是否可以跟机器人组成家庭),但这一段主要是讲人类要不要把机器当成与人类有同等地位的对象,并以人类的伦理去对待机器(比如帮助、同情机器人,机器人可以如何“死亡”等)。

利他行为是一种道德行为。那么,如果利他行为是本文的重点,作者应重点解释考察人类对机器人的利他行为的重要性,为什么这种行为可以代表人类对机器人的道德性(比如,在人类社会中,是否利他行为是最基础的道德行为,因此需要通过考察对机器人的利他行为来体现人类是否赋予了机器人被道德相待的地位)。如果本文的重点是人类是否将机器人视为有主体性的、与自己平等的“物种”,利他行为只是对平等“客体”的行为表现之一,这 downstream

的结果之一，那就应对研究这种心智认知的重要性加以解释，并 justify 为什么要选择利他行为，而不是其他行为(如合作、情感归属等)。

作者回复：感谢专家的详细意见。“道德考量”(moral consideration)是指一定程度的道德关怀，并同时得到一些基本的福利保障(Lund et al., 2007)，当个体将人或物纳入了道德考量的范围也就将其纳入了道德圈的范围。“道德地位”(moral status)我们在文献中并没有看到有具体的定义，这可能是因为在提出“机器人是否拥有道德地位”这一问题的讨论主要集中在伦理学领域，在伦理学的文献中并未定义，因此我们在文章中进行了补充，认为道德地位与道德考量的含义相似，当人们将机器人赋予道德地位后也就是将其纳入道德考量的范围，已有的道德圈文献已证明个体更有可能对道德圈内的个体做出利他行为。为避免“实体”这一名词引起的混淆，我们在文章中将其更正为“人或物”。此外，是否会对机器人做出道德行为是人工智能伦理中一大热门问题，对机器人做出道德行为的发出者是人类，在第一段中我们也提到人类“从祖先那里继承的道德理论与道德实践已经与快速更新迭代的科技生活脱节，无法满足当代社会所出现的新的道德问题”，所以无法去回应是否应当对机器人做出道德行为这一伦理问题。

我们选择对机器人的利他行为是因为这一行为最能代表个体将机器人纳入了道德考量的范围，因为个体会保障道德考量范围内的人或物的福祉和应有的道德权利，这也是我们在文章中举的例子，有国外的工程师愿意帮助机器人打官司保障机器人的利益。

(2) 多元文化经历的提出

作者在第四段提出多元文化经历时，逻辑是，机器人对人类而言是新颖事物，而多元文化经历是影响人们对新颖事物接受程度的一个重要前因。然而，影响人们对新颖事物的因素还有很多，比如人格特质 openness to experience，比如认知 uncertainty avoidance，比如目标导向 learning orientation，甚至马基亚维利主义或许都有可能影响对机器人的态度(一切包括机器人都是自己实现目的的手段)。所以，作者应该解释为什么要选择从文化经历的视角来研究对机器人的开放性态度。另外，第二三段是讲是否给予机器人道德地位，这一段是讲是否接受新事物。赋予机器人道德地位和是否把机器人当一个新事物来接受并不是同一件事，人类可以接受机器人这一新事物，但或许只是将其视做其他技术一样的工具，而不会给予其与人类平等的道德地位。

作者回复：非常感谢专家的宝贵意见。的确在多元文化经历的提出上存在逻辑漏洞。因此我们修改了提出的思路，从道德角度出发，根据已有文献中证明的道德判断是更灵活的过程，而多元文化经历的一大优势就是认知灵活性这一角度出发引出了我们的自变量。我们将以上观点整合进 1.3 部分“多元文化经历促进人类对机器人利他行为的心理机制”，新增部分在文中已标红，请专家再审阅。

(3) 多元文化经历与对机器人态度之间的积极关系

作者提到了多元文化经历是双刃剑，并简要回顾了其负面及正面影响，接着便提出多元文化经历对人类对机器人的态度会是正面影响，这个逻辑是什么呢？既然多元文化经历也有负面的影响，那这些负面机制为什么就不会导致人们对机器人持排斥或不道德的态度呢？正如已有研究所示，多元文化经历带来的相对道德主义可能让人们觉得机器是可以被 abuse 和 slave 的；比如多元文化适应不佳的个体，可能更觉得机器人是个冲击和威胁，挑战了人类的权威和道德地位，将其视为外群体并对其持排斥态度。因此，作者认为多元经历这把双刃剑对机

机器人而言只是个正向的“单刃剑”的原因是什么呢？是否可以探讨什么条件下，多元文化经历会使人们对机器人产生积极态度，什么条件下产生消极态度呢？

作者回复：谢谢专家的宝贵意见。由于我们换了引入多元文化经历的角度，从多元文化经历的优势认知灵活性出发引出了这一变量，因此虽然后文综述了多元文化经历的双刃剑效应，但主要研究问题还是集中在多元文化的积极后果上。关于专家提到的“什么条件下，多元文化经历会使人们对机器人产生积极态度，什么条件下产生消极态度”这一问题本研究虽然未做探讨，但是在讨论中我们将其放入了未来展望的部分。

(4) 多元文化经历与对机器人的心智知觉之间的关系

虽然作者论述了多元文化经历可以提升人们的认知灵活性，拓宽了想像力和对新事物的包容度，但并未解释为何这种接受与包容会体现在对机器人的心智感知上。例如，多元文化经历的人对世界及事物的多元性有更丰富的感知，对变化与不确定性有更强的适应性和安全感，因此，或许觉得机器人这种新异事物是不可怕的。但为什么这种认知灵活性会让人类觉得机器人是具有知觉的道德主体呢？为什么会赋予他人格地位呢？这中间有一个逻辑 gap。

作者回复：感谢专家的宝贵意见。如专家所言，多元文化经历较高的个体会对事物有更丰富的感知和认知灵活性。这些更丰富的感知与灵活性同样会体现在对机器人的心智感知上，尤其是面对机器人需要帮助的情境时更能够激发对机器人的心智感知。而以往研究也表明，增加对机器人的心智感知则会进一步赋予机器人人格地位，并增加对机器人的利他行为(Tanibe et al., 2017)。另外，根据以往研究，多元文化习得促使个体更加强调关爱、公正和平等主义等保护个体权益的价值观(Hu et al., 2020)，进而会更多地增加知觉到的机器人道德权益，使人与机器人的关系更加平等。需要说明的是，我们认为多元文化经历可以提升机器人的“主体(agent)”地位，但此主体并非道德主体(moral agent)，而是指机器人有资格成为有心智的主体并能够获得道德权益。

我们将以上观点整合进 1.4 部分“多元文化经历促进人类对机器人利他行为的心理机制——心智感知的中介作用”，新增部分在文中已标红，请专家再审阅。

(5) 心智知觉与利他行为之间的关系

关于“心智知觉”这一段的论述非常重要，有助于建立心智知觉与对机器人利他行为之间的逻辑关系。根据作者的阐述，心智知觉中的经验性会赋予对方道德权益，即认为对方应该被道德地对待；而主体性赋予对方道德责任，即对方也是在道德上应该承担责任并对结果负责的主体。那么是否相对于主体性而言，经验性更加决定了人们对机器人的利他行为呢？因为利他行为是对机器人的帮助，是在认可机器人所享有的道德权益。还是说经验性和主体性一样重要呢？那么二者是如何分别促进了人们对机器人的利他行为呢？

作者回复：感谢专家的宝贵意见。在人类对机器人实施利他行为时，人类的角色是道德主体，而机器人的角色是道德客体，所以感知机器人体验性的增加会增强人类对机器人的利他行为。我们修改了假设 2 的表述，即多元文化经历通过提高对机器人的心智知觉而提高对机器人的帮助行为。具体而言，相对于能动性，体验性更加决定了人们对机器人的利他行为。

(6) 东西方文化的差异的作用

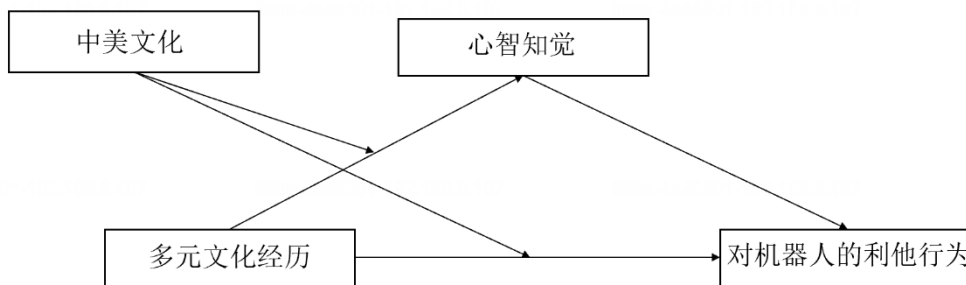
作者提出“希望探讨”不同文化背景下的人机关系与感知，那就应该有理论、过往研究和自己的 theoretical arguments 来支持；作者在图 2 中将东西方文化描述为一个调节变量，就应该有具体的假设。比如东方文化背景的人，其多元文化经历对机器人心智知觉及对机器人

利他行为会产生什么影响？西方文化背景的人呢？二者的差异是怎样的？其原因何在？

同时，为什么东西方文化对多元文化经历→心智知觉的关系没有调节作用，而只对利他行为有调节作用呢？为什么没有调节多元文化经历→心智知觉→对机器人利他行为的链条呢？为什么整个理论模型不可以描述成一个 mediated moderation 呢？

作者回复：感谢专家的宝贵意见。我们认为多元文化经历对机器人利他行为的影响可能存在文化差异。首先，机器人的概念和想法源自西方，最早可以追溯到古希腊时期。所以机器人本身从象征符号层面就有浓厚的西方文化意蕴。所以我们推测多元文化经历对于提升中国人对机器人的心智知觉要高于西方人。另一方面，不同文化下的人们对机器人的接受度存在差异。例如有研究发现，相比于西方国家的民众(如西方)，东亚国家的民众(如日本)对机器人的接受度更低。而对于机器人的接受程度本身也会影响到对机器人的利他行为，所以我们预测经历更多的多元文化经历的中国人可能会比西方人更能增加对机器人的利他行为。据此，我们提出假设 3：东西方文化背景能够调节多元文化经历与对机器人利他行为和心智知觉的关系。具体而言，中西文化调节了直接路径(假设 1)以及心智知觉的中介作用的前半段(假设 2)。理论模型图如下图所示。

我们将以上观点补充进了“1.5 中西文化的调节作用”这部分，并将新增文字标红，请专家再审阅。



2. 论据的准确性

作者在第三段提到，影响对机器人做出利他行为的因素有两类。一是机器人自身的物理和行为特征，二是被试的特征。在对后者举例时引用了 Martin et al (2020)，但该研究看的还是机器人的行为特征(是故意还是无间掉了东西)是否会影响孩子们的帮助行为，并非被试特征。

作者回复：感谢专家指出此问题。之前引用的文献的确不妥。我们重新从被试的拟人化倾向和共情倾向来佐证被试的特征对机器人利他行为的影响。新增加的论据如下：

例如，拟人化倾向(指将人类的外部特征与内在心理状态赋予机器人的倾向)越高的个体会对机器人有更多的共情，从而提升对机器人的接受度和利他行为(Kühnlenz et al., 2013; Nielsen, Pfattheicher, et al., 2022)。另外，“诚实-谦逊(Honesty-Humility)”人格倾向较高的个体会在博弈游戏中给 AI 分更多的钱(Nielsen, Thielmann, et al., 2022)。

3. 引言、理论与假设

(1) 建议作者按通行的写作规范，将 1.引言和 2.理论和假设分为两个部分。在引言中强调研究问题的意义、概念的选择、对理论的潜在贡献；在理论和假设中更详尽地 theorize 为何多元文化经历会影响心智知觉，心智知觉为何会影响对机器人的利他行为，以及为何多元文化经历对后两者的影响具有文化差异。同时，提出具体的、有方向预测性的假设。例如：

假设 1: 多元文化经历会提高对机器人的心智知觉。

假设 2: 多元文化经历通过提高对机器人的心智知觉而提高对机器人的帮助行为。

假设 3: 东西方文化背景对多元文化经历与对机器人的心智知觉间的关系(H1)起调节作用, 表现为, 相对于西方文化背景的人, 东方文化背景的人的文化经历对心智知觉的提升作用更强(或更弱)。

假设 4: 东西方文化背景对多元文化经历通过影响对机器人的心智知觉从而影响对机器人利他行为的关系(H2)起调节作用, 表现为, 相对于西方文化背景的人, 在东方文化背景的人群中, 这一正向链条关系更强(或更弱)。

作者回复: 十分感谢专家的建言! 结合下一条意见, 我们将“多元文化经历促进人类对机器人的利他行为”这个主效应作为假设 1。修改后的所有假设如下:

假设 1: 多元文化经历促进人类对机器人的利他行为。

假设 2: 多元文化经历通过提高对机器人的心智知觉而提高对机器人的帮助行为。具体而言, 相对于能动性, 体验性更加决定了人们对机器人的利他行为。

假设 3: 东西方文化背景能够调节多元文化经历与对机器人利他行为和心智知觉的关系。具体而言, 中西文化调节了直接路径(假设 1)以及心智知觉的中介作用的前半段(假设 2)。

(2) 实验一和三的结果显示, 多元文化经历和对机器人的利他行为之间是有正向关系的, 但不一定是由对机器人的心智知觉解释的。作者在总讨论中也强调了主效应, 及其存在的其他解释。所以, 作者还可考虑, 是否需要一个多元文化经历-->对机器人利他行为的假设, 与整体的逻辑顺序是否融洽。

作者回复: 感谢专家的意见。本文确实需要一个针对主效应的假设。我们将“多元文化经历促进人类对机器人的利他行为”作为假设 1。

二、实证问题

1. 实验一

(1) 样本

作者提到能过“探测题”收回 266 份样本, 是什么样的探测题呢? 是问被试是否有过海外经历吗? 不然, 在正常问卷星的被试库里, 去过 1-4 个国家的被试比例不可能这么高。应加以说明。

作者回复: 谢谢专家的宝贵意见。探测题是指注意力检查题, 如“此题请选 x”。实验一的被试都是有海外经历 6 个月以上的被试。在修改中实验一的被试部分着重强调了这一点, 谢谢专家的意见, 一开始说明得不够清楚。

(2) 多元文化经历量表

该量表中有两个题项是“观看国外的电视频道 (如: 电视剧、综艺、纪录片等)”, “在电视观看不同文化的庆典”。看电视在平均年龄为 19-30 岁人群中是鲜有的行为了。虽然问题的重点在于文化经历, 但如果有被试确实看了这些节目, 但是不是通过电视, 他可能也会 frequency 打分低, 因为被试回答问题时关注的焦点可能是看节目的渠道。应该根据时代的变化对题项措辞加以调整。

作者回复: 谢谢专家的宝贵意见。的确现在年轻人很少看电视了, 所以在实验二的量表中我们修改了措辞, 添加了“在网络上观看”这一更为普遍的途径。

(3) Threats to internal validity

a. 变量测量顺序

作者在测量了对机器人的心智知觉后再测量对两个机器人的帮助倾向。这里可能产生 **demanding effect**, 被试可能被心智知觉的问题提醒而觉得机器人可能是有心智的物体, 从而 **inflate** 他们的帮助打分。

作者回复: 谢谢专家的意见, 的确可能存在这样的问题。我们在收实验一数据时除了人口学变量外其余量表的顺序都进行了随机。

b. DV 的引导语及图片

引导语中说“家用机器人 **Doppy** 并不擅长做饭, 每次做饭 **Doppy** 都很苦恼”; “**Doppy** 在做家务时不小心打碎了玻璃杯”。这个引导语, 加上图片, 已经预设了机器的人格化, 比如有人的名字, 有人的感情(苦恼), 有人的行为(“不小心”及抱头等), 被人格化装扮(戴上了围裙)。这会严重影响被试的帮助意愿。即便在之前的心智知觉中打了低分的被试, 这时也可能被语言和图片 **prime** 出比之前较高的心智知觉, 从而体现出与之前心智知觉打分不一致的帮助倾向。

作者回复: 谢谢专家的宝贵意见, 非常有道理。的确在图片的指导语中预设了机器人的心智倾向, 因此在重新收集的实验二的数据中, 我们换掉了图片测量对机器人利他的意愿, 而是改成了量表测量。

(4) 假设检验

PROCESS 模型中是否包括了控制变量, 为什么包括或不包括, 结果有何不同?

作者回复: 谢谢专家的宝贵意见。我们在模型中包括了人口学变量, 因为之前有研究表明这些人口学变量对因变量有影响因此控制掉了。

2. 实验二

(1) 题目

实验二与实验一一样, 都是检验中介关系, 为什么实验一的标题是“是否提高“, 实验二的标题是“为何提高”? 这给人的预期是, 实验一是看多元文化经历-->利他行为, 实验二才是看实验一是看多元文化经历-->心智知觉 --> 利他行为。

作者回复: 谢谢专家的宝贵意见。我们实验一和实验二的描述不够严谨, 已进行更正。实验一的题目为“多元文化经历提高对机器人的利他行为: 心智知觉的中介作用”, 实验二的题目为“多元文化经历影响对机器人利他行为的因果实验”。

(2) 样本特征

未提供跟实验一一样详细的样本人口特征描述。

作者回复: 谢谢专家的宝贵意见。在新的实验二中已进行了修改补充。

(3) 自变量(多元文化经历)的操控

a. 作者提供了多元文化组的操控材料。其他两组的启动材料是怎样的? 是与多元文化一模

一样，也让填入城市名称，并想像文字所描述的经历吗？如果家乡不是城市呢？日常经历组应该怎么填那个空呢？如果最近的一次经历不是城市呢？还是要求了被试必须回想最近一次游览城市的经历呢？

如果另两组与多元文化经历组不同，又是怎样的呢？多元文化组的城市经历明显是积极的、愉悦的、充满好奇和开放心态的、愿意接纳与探索的。如果该组被试更愿意帮助机器人，或许不是由于文化经历的多元性导致的，而是由于这种愉悦且开放的经历本身激活了被试的开放心态。另两组的经历在 *emotional tone and valence, openness* 等方面是 *equivalent* 的吗？

作者回复：谢谢专家的宝贵意见。我们根据前人文献(Lu et al., 2017; Maddux & Galinsky, 2009)修改了启动多元文化经历的材料，要求多元文化经历组回忆“您曾去过的一个外国国家，详细描述一下您在国外生活的那段时间都有哪些生活经历。比如，在那个国家经常发生什么事，您遇到了哪些人，您在国外生活是什么感受，您看到了什么、做了什么，想到了什么”，并写下不少于 100 字的描述。家乡文化经历组的指导语是“请您回忆您的家乡，详细描述一下您在家乡生活的那段时间都有哪些生活经历。比如，在家乡经常发生什么事，您遇到了哪些人，您在家乡生活是什么感受，您看到了什么、做了什么，想到了什么”，日常生活对照组和指导语是“请您回忆并详细描述下您上一次去超市的经历”，两组均被要求写下不少于 100 字的描述。

b. “当被试阅读完启动材料后，要求被试写下 100 字描述其想象到的国外生活经历/家乡生活经历和日常生活。”是读完上面的城市游览文字后再写 100 字吗？其目的是什么呢？

作者回复：谢谢专家的宝贵意见。让被试读完城市/村镇游览的文字后再写 100 字描述是为了加强启动效应。

c. 更重要的问题是，为什么除了多元文化组外，还有两组呢？设计三组的逻辑是什么？在后面，作者将日常生活组称为“控制组”，所以多元文化组和家乡文化组是 *treatment groups* 吗？如此的话，应该把多元和家乡组分别与控制组(即日常生活组)做比较。那家乡 vs. 日常的逻辑是什么呢？家乡文化组的功能是什么呢？作者期望这三组呈现什么样的差异呢？是认为家乡会让人更闭塞，所以比日常组更不会感知机器人有心智吗？还是说家乡与日常都是控制组？那有两个控制组与多元文化组相对照的逻辑是什么呢？所有这些设计对支持假设有什么功能呢？这些都应该在实验二的引言中加以说明。

作者回复：谢谢专家的宝贵意见。家乡和日常都是控制组，家乡是被试熟悉的生活环境，而多元文化经历是新异的体验，但两者都属于文化经历，因此还需要一个不涉及文化体验的日常生活对照组。所以设置了三组条件，在文章中也进行了补充说明。

(4) 操控检验

a. 操控检验是如何进行的作者完全没有描述。

作者回复：谢谢专家的宝贵意见。操纵检验的三道题目来自多元文化经历量表，进行了细微修改，分别为“我和国外文化的人接触很多”“我和国外文化的事物接触很多(比如电影、音乐、书籍等)”“我对探索国外的文化风格很有兴趣”。已在文章进行了补充。

b. 操控检验放在心智知觉与利他行为之前，是否会 *remind* 被试假设，从而 *interfere* 后面的

回答呢？为何没考虑做单独的 pilot test 来检验操控，并排除任何其他可能的 confoundings？

作者回复：谢谢专家的宝贵意见。在新的实验二中，操纵检验与心智知觉和利他量表的呈现顺序是随机的。

c. 作者汇报了 omnibus F-test 的结果，并未汇报三级之间 pairwise comparison 的结果。这跟 (3)c 中提到的问题是相关的，不了解设计三组的逻辑，也无从知道两两比较希望是何结果。

作者回复：谢谢专家的宝贵意见。在修改中我们重新汇报了新实验二的事后比较结果。

(5) 心智知觉及利他行为的分析

也只是汇报了 omnibus F-test，并未检验两两比较的显著性。

作者回复：谢谢专家的宝贵意见。这里我们重新汇报了新实验二的事后比较结果。

(6) 中介效应

作者未汇报中介模型的检验，这或许是因为 X 到 M 就不显著(X 到 Y 不显著是检验 indirect effect 的必要条件)。但正如实验二标题和图 1 假设一所示，要检验的核心是中介效应，作者应该对未做/无法检验有所交待。

作者回复：谢谢专家的宝贵意见。在新实验二中我们进行了中介效应的检验。

(7) 其他变量

作者在实验二的引言和结果讨论，以及总讨论中都提到，除心智知觉外，还加入了“合作意愿”和“道德宽容”，甚至在总讨论中反复提到这两个变量上的结果，并进行了大量讨论。可见作者除了测量心智知觉外，还测量了合作意愿及道德宽容。这两个变量是与利他行为平行的、反映“道德考量”的 DV 吗？也受心智知觉的影响吗？还是与心智知觉平行的中介变量呢？所有这些完全没有在理论与假设中提出，也没有在 S2 的变量测量中提及。

作者回复：非常感谢专家的宝贵意见。之前测量合作意愿和道德宽容是希望探索多元文化经历可能引发的多种积极后果。为避免引起迷惑我们在新的实验二中去掉了相关变量的测量也不再在文章中进行汇报。

(8) 作者提到可能是由于在线研究，导致启动效应不强。如果条件许可，作者可否考虑在实验室重新开展该研究，甚至可以使用真实的机器人，看被试的帮助行为。比如给 Experimenter 安排一个机器人“助手”，貌似“它”的材料撒在地方，看被试是否会帮助收拾。当然这只是 intuitive 的想法，作者可设计其他更好操作的实验。

作者回复：谢谢专家的宝贵意见。我们重新设计了实验二，一开始有考虑实验二在实验室中进行，使用真实的机器人，但是由于疫情原因导致线下实验室开放时间不定且实验室未有机器人的相关设备，因此新实验二还是使用了线上实验。

3. 实验三

(1) 假设 1 分析

应该先分析假设 1(或我上面列出的假设 1、2)在西方文化中是否成立。这个信息是关键且有必要的,可以检验同一个因果链是否在另一个文化中存在。这也是下一步比较文化差异的基础。

作者回复: 谢谢专家的宝贵意见。在实验三的修订中我们首先检验了在西方被试中假设 1、2 是否成立。

(2) 文化差异分析

a. 作者提到了两个控制变量(性别、关系类型)。首先,“关系类型”是什么?其次,在跨文化比较的研究中,两个样本往往在诸多方面存在差异,且可能对目标变量及关系产生影响,因此应该将这类变量加以控制。例如在人口统计变量上的差异。据作者对实验一和三的样本描述,被试在海外经历上具有巨大差异,却未加控制。另外,或许中西文化下的一般性帮助意愿本就存在差异,比如帮助陌生人(或许西方人更会帮助)、帮助没做好事的孩子(或许中国人更愿意帮助)。这些一般帮助倾向未加控制,会严重影响对文化差异的分析与解读。此外,正如作者在讨论中提到的,已经有研究发现中国人对机器人的负面态度明显高于西方,这种已经有的结论和变量更应该做为控制变量去测量。

作者回复: 谢谢专家的宝贵意见。“关系类型”是之前测量其他变量时的条件,在新修订的实验三中已经去掉了。在实验三中,我们控制了人口学变量,性别、年龄、最高学历、主观社会经济地位,但由于中西两国经济状况存在客观差异,且研究一的结果表明客观家庭收入与自变量、因变量和中介变量均不相关,故未将客观家庭收入纳入分析。我们也将海外经历、对机器人的负面态度纳入了控制变量。此外,中西文化在一般帮助意愿上可能存在差异,这里的确没有进行基线水平的测量,在讨论部分进行了说明。

b. 不论控制了哪些变量,也应该在回归结果中汇报包括与不包括控制变量模型结果,说明结果是否有显著差异。

作者回复: 谢谢专家的宝贵意见。我们重新修订了结果报告部分,汇报了包括控制和不包括的模型。

c. 虽然交互效应不显著,但两个主效应是显著的。文化背景的主效应,由于没有注明中国与西方文化的 code,读者无法解读回归系数;在讨论中作者提到西方被试比中国被试的平均利他行为意愿要高,应该补充上中国和西方被试的平均利他行为打分(estimated means)。

作者回复: 谢谢专家的宝贵意见。我们在修订过程中对回归系数和平均利他行为意愿进行了汇报。

4. 综合问题

(1) 虽然在自检报告中,作者汇报了三个研究的计划样本量为 300 以及有效的样本量。但选取 300 的依据是什么?可曾做 power analysis?

作者回复: 谢谢专家的宝贵意见。我们之前用 g-power 做过被试量的预测,自检报告重新进行了修订。

(2) 三个研究的 correlation table 都应该汇报所有变量(包括控制变量)的 mean, sd, &

correlations.

作者回复：谢谢专家的宝贵意见。在修订中三个研究都汇报了所有变量的均值、标准差和相关系数。

(3) 总讨论中说“实验 3 在西方样本中重复了实验一的发现，验证了多元文化经历对机器人的利他行为的提升作用”，但作者是把实验一和三两个样本放在一起后检验的该效应，并未汇报单独在西方样本中的检验结果。

作者回复：谢谢专家的宝贵意见。我们重新汇报了在西方样本中的检验结果。

三、其他问题

1. Typo, p.7, line 4, “当我们讨论如果对待回应机器人时”应该是“如何”。

作者回复：谢谢专家的意见。已修改，对全文的错别字进行了修订。

2. “5.结论”略嫌 over-interpretation.

作者回复：谢谢专家的意见。我们重新修改了结论部分，并将其标红。

.....

审稿人 2:

论文考察了多元文化经历对人类对机器人的利他行为的作用，特别是提出了心智知觉的中介作用。选题具有一定的理论价值，但是目前论文从理论基础、假设提出、实验设计、结果汇到等方面都存在较大的问题，需要进一步大幅提升。具体建议如下：

1.对相关理论背景和文献的梳理不够系统，显得较为零碎。缺少明确的假设提出部分。从多元文化经历-到心智直觉—再到利他行为的每个逻辑链条的理论推导或逻辑论述都不充分，需要进一步的加强。特别是，为什么多元文化经历会让人们感觉机器人更有心智(更像人?) 这个关键的逻辑点在文中并没有阐述清楚。

作者回复：感谢专家的意见。我们在修改稿中更为详细地阐述了“多元文化经历-心智知觉”、“心智知觉-利他行为”的中介链条。针对“为什么多元文化经历会让人们感觉机器人更有心智(更像人?)”这个问题，我们认为多元文化经历的人有更高的创造力和认知灵活性，对世界及事物的多元性有更丰富的感知，对变化与不确定性有更强的适应性和安全感。这些更丰富的感知与灵活性同样会体现在对机器人的心智感知上，尤其是面对机器人需要帮助的情境时更能够激发对机器人的心智感知。而以往研究也表明，增加对机器人的心智感知则会进一步赋予机器人人格地位，并增加对机器人的利他行为(Tanibe et al., 2017)。另外，根据以往研究，多元文化习得促使个体更加强调关爱、公正和平等主义等保护个体权益的价值观(Hu et al., 2020)，进而会更多地增加知觉到的机器人道德权益，使人与机器人的关系更加平等。

我们将以上观点整合进 1.4 部分“多元文化经历促进人类对机器人利他行为的心理机制——心智感知的中介作用”，新增部分在文中已标红，请专家再审阅。

2 .多元文化经历对针对机器人的利他行为的主效应本身还需要在其他实验情境以及 field study 等情况下做出更充分的检验，以此来验证其可靠性。

作者回复：感谢专家的意见。我们重新设计并实施了实验 2，旨在验证多元文化经历对机器人利他行为的主效应，并验证心智知觉在其中的中介作用。

3.除了心智知觉之外，是否存在其他可替代性的中介变量，能够解释多元文化经历的作用？例如，是否更友善、更有共情能力？不仅对机器人，对人类也会展现出更多的利他行为。需要再深入文献梳理出可能的替代性解释，并予以排除。

作者回复：感谢专家的意见。我们在新实施的实验 2 中补充了积极与消极情绪作为协变量，进而排除情绪对主效应的干扰作用。针对专家提出的友善、共情等因素，我们在后文的总讨论部分论述了以上因素或许可以解释多元文化经历对机器人利他行为的作用，并呼吁未来的研究进一步检验这些可能性。新增加的部分文字如下：

本研究虽然排除了积极和消极情绪对实验结果的影响，但还有一些变量，例如友善与共情能力，可能对实验的主效应起到一定的干扰作用。未来研究应探讨除心智知觉外其他可替代性的中介变量，使实验结果更加稳健。

4.实验二只得到方向性的结果，并没有达到统计显著。对多元文化经历的操纵不是很严谨，是否进行了前测或操纵检验，是否参照了已有文献的方法？需要对实验的刺激物、流程、测量方式等进一步的完善。结果汇报过于简略，应对照期刊近期发表的文章来提升。

作者回复：谢谢专家非常仔细的意见。我们之前实验二的启动范式的确不够严谨，因此新实验二参照前人文献(Lu et al., 2017; Maddux & Galinsky, 2009)进行启动。此外，为确保能成功启动，我们与前人文献保持一致，所招募的被试全是具有海外经历的被试，在新实验二中成功启动了多元文化经历，多元文化经历组的被试在操纵检验上的得分显著更高。我们将实验二全部修改了，对流程和测量方式也进一步完善，请专家再审阅。

5.实验三的目的是检验中西差异，但是文中只提到外国，并没有清楚地交代中西文化背景是如何测量或区分的。实验设计不够严谨，实验结果汇报混乱。统计结果和作者自陈的研究结论矛盾。

作者回复：谢谢专家的宝贵意见。由于之前论述不够严谨，我们在 Mturk 上收集的被试不能全部归为西方被试，实验三的目的也是想探讨实验一的结果能否推论到其他西方文化中。此外，我们把整篇文章的实验部分全部重写了，细致梳理了实验结果和结论的汇报，请专家再审阅。

第二轮

审稿专家 1：

感谢有机会再次审阅《多元文化经历提升人类对机器人的利他行为：心智知觉的中介作用》，也感谢作者对上一轮提出的问题作出了详细的回答。作者改进了理论阐释与逻辑，重新设计了实验二并修订了实验三。下面对依然存在的问题、有些问题的解决思路等加以描述，希望对作者有所帮助。

一、理论与逻辑

1. 这一版对道德相关的概念做了更清楚地解释，说明了机器人拥有道德地位的意义，为什么心智感知是将一个对象纳入道德考量的前提，以及为什么帮助行为是对道德考量对象的重要行为标志。然而，理论构建的难点仍然在于多元文化经历为何能提高心智感知。作者提出

认知灵活性是主要机制。然而，multicultural experience research 中发现 cognitive flexibility 有可能带来积极和消极的结果。一方面，认知灵活性是创新的基础，例如接纳新的想法、突破既有的思维和做法；另一方面，认知灵活性也可能是不道德行为的前提，使得人们破坏规则、“另辟蹊径”如欺骗等。还有研究发现多元文化经历让人更 moral relativism, 更加使用不同标准，而不是平等的、universal 的标准去做道德判断，从而为自己的不道德行为找到 idiosyncratic 的理由或情境化的借口。因此，不论是从多元文化本身的研究看，还是从认知灵活性看，对机器人心智感知与道德对待(i.e., 帮助)的影响都不明确。个人有以下建议。

(1) 加强说明为什么多元文化经历带来的 cognitive flexibility 更容易在机器人身上发生积极效应。也就是，多元文化提高了认知灵活性，从而提高对新事物的认可、接受和包容，从而更可能接受机器人这一新技术产物；但接下来，为什么对这一新事物的不排斥会升级到认为其有心智并需要提供道德考量。回答这个问题有助于打通多元文化经历—心智感知—对机器人利他的逻辑链条。

作者回复：谢谢专家的宝贵意见，对我们启发良多。根据专家的建议，我们重新思考了多元文化经历提升对机器人利他行为可能的原因，除了认知灵活性外，多元文化经历打破了对外群体的偏见和歧视，而机器人于人类来讲属于外群体(Smith, Šabanovi, & Fraune, 2020)，此外，认知灵活性可以让我们从不同的角度看待问题，提升了人们的共情水平和理解范围(Park & Yu, 2017)，综合上述原因，我们认为多元文化经历很有可能提升对机器人的利他行为，相关的表述也在文中标红，请专家审阅。

(2) 另一个思路是不要局限于 cognitive flexibility, 思考多元文化经历—心智感知之间建立关系的更多可能。作者在文中提到了多元文化的多种影响，如“多元文化经历可能会减少偏见和歧视(Tadmor et al., 2012)，增加个体对他人的广义信任(Cao et al., 2014)，提高创造力(Leung et al., 2008)和更能容忍不确定性(Adair & Xiong, 2018)”“对世界及事物的多元性有更丰富的感知，对变化与不确定性有更强的适应性和安全感”“多元文化习得促使个体更加强调关爱、公正和平等主义等保护个体权益的价值观”，这些研究发现是否更有助于说明为什么多元文化经历可能增加人们对机器的道德考量。例如，偏见和歧视的本质是以刻板的特征对某个群体产生不公正的判断和行为、甚至 inhumane 的行为；而对机器人有道德考量，需要个体摆脱“机器人是机器”群体的一员的认知，不是用对机器的刻板印象和态度来对待机器人。再比如，广义信任意味着人们在更广泛的情境中具有较高的一般性信任感，放到机器人情境下，广义信任是否有助于人类将智能机器不当做威胁而对其持有友善平等的态度？创造力和容忍不确定性都包括对新事物的接纳、尊重和使用。不过，这里的逻辑难点依然在于，由于多元文化经历可降低偏见与刻板思维、提升信任与包容等，从而提升了对机器人的接纳后，下一步如何升级为对机器人的心智感知。建议作者进一步深挖，多元文化的 ABC 个后果中，机器人心智感知的 OPQ 个前因中(如作者提到的拟人化倾向、共情、诚实-谦逊人格)，是否 A 与 P、B 与 O 之间可以建立关联。

作者回复：十分感谢专家提供的思路！按照专家的建议，我们尝试将多元文化经历导致的多种后果(如减少偏见、增加信任、提升创造力等)与心智感知的若干前因变量之间建立联系，从而更有逻辑地论述两者之间的关系。我们在 1.4 部分增加了如下文字，在文中也已经标红。

另一方面，多元文化经历会减少对外群体的偏见与歧视，更少以刻板的特征对某个群体进行不公正的判断。因此，多元文化经历更丰富的个体可能更能摆脱“机器人只能是机器”的刻板印象，并减少对机器人“非人化(dehumanization, 即对人类基本特征的否定)”倾向，从而更倾向于将机器人纳入内群体并赋予机器人以人的特性。根据心智感知理论，一旦人类对机器人的感知相似度(perceived similarity)增加，并产生试图与机器人建立社会联结(social

connection)的动机时,便会触发心智感知(Waytz et al., 2010),即认为机器人更加富有思想与情感。再者,多元文化经历更丰富的个体在更广泛的情境中具有较高的广义信任,在机器人情境下,这种广义信任可能使得人类不再倾向将智能机器人视为威胁而对其持有更友善平等的态度。

(3) 第三个思路比较突破现在框架,不一定可行,仅供作者开拓思路。可否将多元文化经历是否会提升对机器人的利他行为 Frame 成一个开放问题,把心智感知当做边界条件,即只有当拥有多元文化经历的人同时对机器人有心智感知时,他们才会对机器“人”表现出更多的“人际”关怀行为。基本 story 依然是 AI、技术等的发展带来的对机器人的伦理考量问题,已有研究发现机器人自身特征(如外形、声音)可能增加人们对其的帮助行为,但较少探索个体特征的作用。本文提出个体多元文化经历这一因素,因为大量研究显示多元文化经历会提升人们对新事物的接纳程度、信任等等..., 这些都可能增加这类人群对机器人这一新技术产物的接纳程度;然而是否将机器人真正当人来对待,并向人类互助一样对机器人做出帮助等利他行为呢?关于道德考量的研究发现,一个重要的条件是对该对象的心智感知。因此,只有当既有多元文化又有心智感知时(但这两者是 orthogonal 的),才可能去像帮人一样帮机器人。这样作者就不必建立多元文化与心智感知之间的关系,而把心智感知做为一个独立的条件,是促使多元文化经历对机器人的积极影响上升到人际伦理的一个 independent condition. 当然,这些还需要看已有的数据是否支持。

作者回复:十分感谢专家提供的思路!如专家所言,这对现有研究框架突破性较大,我们可能还要重新思考中介机制并设计新实验验证之,这在有限的规定修稿时间内可能不允许。其次,在重新进行数据分析后,结果表明心智感知作为调节变量在三个研究中都不成立。使用 SPSS 26.0 的 PROCESS 插件(Hayes, 2013)中的模型 1 进行分析,研究 1a 中不包含控制变量时调节作用不显著($p = 0.31$),包含控制变量时调节作用也不显著($p = 0.30$);研究 1b 中不包含控制变量时调节作用显著($p = 0.02$),包含控制变量时调节作用不显著($p = 0.06$);研究 2 中调节作用不显著($p = 0.59$)。综合以上考量,我们还是倾向于将心智感知作为中介变量来构建整体的研究框架,并更为详细地解释了将心智感知作为多元文化经历的后果与利他行为前因的理由(亦可见上条回复)。不过我们还是非常感谢专家为研究框架提供的建言!

2. 作者强调了机器人的体验性心智才能提升人类对其的帮助,而能动性心智只会使人觉得机器人也负有道德责任。但作者的假设是一般性的、总体的心智感知。然而在提出假设之后又加了一句“具体而言,相对于能动性,体验性更加决定了人们对机器人的利他行为。”这样比较混乱。作者要么在 theory 中不做类别区分,只说明觉得客体有心智时,人们为什么会去帮助。等到某个具体 study 要进行区分时再在那里的小引言中说明,该项研究目的之一是进一步区分文献中提出的两种心智感知。要么就是所有的实证与理论都保持一致,不做区分,或都做区分/只聚焦于体验性心智。

作者回复:感谢专家的宝贵意见。我们按照专家的建言,在前言中不再区分心智感知感知的具体维度,将假设中的“具体而言,相对于能动性,体验性更加决定了人们对机器人的利他行为。”删除,避免混淆。同时在研究 2 的前言中对心智感知的两个维度进行区分。

另外,研究一并未区分心智感知的具体维度,即能动性与体验性,无法更深入了解心智感知的分维度对主效应的不同解释机制。我们推测,在人类对机器人实施利他行为时,由于人类的角色是道德主体,而机器人的角色是道德客体,所以机器人体验性的增加会使人类对机器人有更多的利他行为。也有研究发现当人们对机器人做出帮助行为时,会认为机器人有更多体验快乐的心智属性(Tanibe et al., 2017),即体验性的感知。所以,相对于能动性,体验性可能更加决定了人们对机器人的利他行为。

3. 中西方文化的调节作用。感觉这一部分较薄弱，论述中感觉是在说中西方文化对心智感知的主效应；而其实要 argue 的是，为什么多元文化经历—>心智感知这一关系在东方或西方情境下更强或更弱。为什么当中国人和西方人经历了一样的多元文化时，中国人就会对机器人有更强的心智感知？这个逻辑没说清楚。如果作者认为这一链条关系在东西方有很大的差异，且这个差异是有意义的，那就应加强文化的调节作用的论述；如果只是表现程度有所差异，但研究重点还是建立这一链条，而不是文化差异，那建议不要这个调节变量了。

作者回复：非常感谢专家的建议。我们在处理中西方文化的调节作用时确实存在论证不充分的问题。根据专家的建议，我们将重心放在了建立“多元文化经历-心智知觉-对机器人的利他行为”这一模型的跨文化普适性上，去掉了中西方的调节作用。

二、方法

1. 研究二的实验操纵基本上失败，因此 main study 里也没有得到与假设一致的结论。建议在这次结果及已有文献基础上，进一步改进操控，最好做个 pilot study 以确保操纵成功且没有带来 confounding factors 的变化。

作者回复：谢谢专家的宝贵意见。我们通过前人文献中的范式，在操纵检验阶段启动了多元文化经历(以组别为自变量，操纵检验为因变量的单因素方差分析的结果表明操纵成功， $F(2,246) = 3.65, p = 0.03, \eta^2p = 0.29$)，因此没有再进行 pilot study。

2. 实验分析中不应该有控制变量，除非组间在控制变量上显著差异，但这一一般在 random assignment 情况下是不太会发生的。

作者回复：谢谢专家的宝贵意见。因为控制变量在三组之间的差异不显著，所以根据专家的检验，我们去掉了控制变量。

3. 研究三不应用中国样本中的相关系数显著性做为依据来决定美国样本中是否收集相关数据，只要理论上或以往在不同样本中的实证都发现某些变量需要被控制，就应该控制。

作者回复：感谢专家的宝贵意见。我们在研究三里的表述的确不严谨，因此去掉了相关表述，只控制了理论上可能会有影响的变量。

4. 研究三并未发现文化的调节作用，即未发现多元文化经历-心智知觉-帮助行为这一链条在中美有差异。如理论部分所建议的，如果文化差异不是本研究的重点，或是在逻辑上很难辨析文化差异并做出合理假设，建议作者将研究三与研究一并列成研究 1a 和 1b，只是为了在不同文化中验证同一套假设，提高结论的 external validity. 如果还是想看文化差异，可以加一个“additoinal analysis”把文化当个调节变量。

作者回复：感谢专家的宝贵意见。我们将研究一和研究三并成了研究 1a 和 1b，将重心放在了检验中介模型的可推广性上，在 additional analysis 中把文化作为了调节变量。

5. 汇报规范：

(1) Design and proedure 应该在 measures 描述之前

(2) 在文字和表格中都不应使用 \pm ，而是直接汇报 mean, SD.

作者回复：感谢专家的宝贵意见，我们在正文中进行了修改，请专家审阅。

.....

审稿专家 3：

本文探讨了多元文化经历是否影响对机器人的利他行为，以及心智知觉的中介效应。通过三个研究，本文获得了初步的支持性证据，相关关系的结果稳健，因果关系的结果，需要

未来研究进一步重复验证。研究结果具有重要的理论和实践意义。但是，本文也有一些缺陷，需要作出修改。

文章的一个主要问题，是文献综述、研究方法、统计分析、以及讨论的表述不清晰，需要修改补充和润色。

1.“1.1 机器人是否可能拥有道德地位？”——在这一部分中，需要聚焦于讨论机器人的道德地位问题。但是，在这一部分，作者还讨论了人类对机器人的利他行为：“在日常生活中，我们也可能对机器人做出一些微小的利他行为，比如避开上菜机器人的行进路线，帮助扫地机器人处理其无法触及的区域等。”这些讨论应该移到 1.2 部分。

作者回复：感谢专家的宝贵意见。在 1.1 部分放入这些讨论的确不太合适，因此已将其挪至 1.2 部分，在正文中进行了标红，请专家审阅。

2. 作者需避免过长句子的表达，比如：“有心理学家基于“计算机是社会行动者”(computers are social actors)框架(Nass & Moon, 2000)，提出人类对机器的行为是亲社会的，当人们以亲社会的方式对待人类和机器时，类似的动机和认知过程也发挥了作用，其中包括将机器视为人类，将机器进行社会分类，受到机器人的社会影响，以及对机器体验社会情绪(Nielsen, Pfattheicher, & Keijsers, 2021)”。——读者很难理解过长句子所表达的重点，以及与其它句子的逻辑衔接。

作者回复：感谢专家的宝贵意见，我们对文中的所有长句都进行了修改，尽量使其具有可读性。

3. 注意避免错别字，比如““人们对机器人所做的道德行为将直接反应人们是否愿意“——反应，应该是“反映”。

作者回复：感谢专家的宝贵意见，我们对全文又进行了修改，避开了错别字。

4. 引用文献需要为论点服务，比如，在“1.2 人类对机器人利他行为的影响因素”这一部分中，“已有研究表明，机器人自身的物理和行为特征会影响人们对其利他的态度以及之后的行为(Somanader et al., 2011)。当机器人表现得更具自主性(autonomy)时，相比那些人为遥控的机器人，被试会更快地响应机器人的求助请求(Srinivasan & Takayama, 2016)，并且人们会认为遥控机器人的智商更低(Tozadore et al., 2017)。而当机器人展现出更高的智力水平和宜人性时，人们会显著犹豫更长的时间来关掉机器人(Bartneck et al., 2007)。此外，当机器人的拟人化程度越高时，人们会赋予机器人更多的情感特征，并减少在道德两难情境中牺牲机器人的意图(Nijssen et al., 2019)。”——人们认为机器人智商低、关掉机器人、减少牺牲机器人的意图，这些论据并不能说明人类对机器人是否有利他行为，可能反映的是人们对机器人的态度。

作者回复：感谢专家的宝贵意见。我们重新整合了此部分的论据，将“人们认为机器人智商低、关掉机器人、减少牺牲机器人的意图”这部分改为直接反映对机器人利他行为的论据。修改后的文字如下，在文中也已标记为红色，请专家再审阅。

已有研究表明，机器人自身的物理和行为特征会影响人们对其的利他态度以及之后的行为。当机器人表现得更具自主性(autonomy)时，相比那些人为遥控的机器人，被试会更快地响应机器人的求助请求(Somanader et al., 2011; Srinivasan & Takayama, 2016)。Lee 和 Liang(2016)发现，如果机器人在之前的问答游戏中做出了有用的帮助行为，那么被试会更愿意帮助它，说明对机器人的利他行为也会遵循互惠原则。此外，机器人自身的社会类别线索也会影响人们对其的利他行为。Siegel 等人(2009)的研究表明，与发出男性声音机器人相比，男性对发出女性声音的机器人有更多的捐赠行为。

5. “1.3 多元文化经历促进人类对机器人利他行为”，在这一部分中，作者介绍的文献，大多是论证多元文化经历对道德标准的影响。但是，本文探讨的是对机器人的“利他行为”，与道德标准并不相同。因而，这部分应该加强“多元文化经历如何影响利他行为”的论述。

作者回复：谢谢专家的宝贵意见。我们在“多元文化经历如何影响利他行为”这部分的论述的确不够充分，再次查阅文献后，我们重新思考了多元文化经历提升对机器人利他行为可能的原因，除了认知灵活性外，多元文化经历打破了对外群体的偏见和歧视，而机器人于人类而言属于外群体(Smith, Šabanovi, & Fraune, 2020)，此外，认知灵活性可以让我们从不同的角度看待问题，提升了人们的共情水平和理解范围(Park & Yu, 2017)，综合上述原因，我们认为多元文化经历很有可能提升对机器人的利他行为，相关的表述也在文中标红，请专家审阅。

6. 注意写作逻辑：“相比于西方国家的民众(如美国)，东亚国家的民众(如日本)对机器人的接受度更低(Sakuma et al., 2019)。而对于机器人的接受程度本身也会影响到对机器人的利他行为(Nielsen, Thielmann, et al., 2022)，所以我们预测经历更多的多元文化经历的中国人可能会比西方人更能增加对机器人的利他行为”——按照作者所述，若东亚人对机器人的接受度更低，那么，他们对机器人的利他行为应该更少。即使东亚人有多元文化经历，他们对西方文化的接受程度也通常不会高于西方人。

作者回复：谢谢专家的宝贵意见，在文化的调节作用这一部分，我们表述的逻辑不够清晰，在逻辑上很难辨析文化差异并做出合理假设，因此我们将重心放在了建立“多元文化经历-心智知觉-对机器人的利他行为”这一模型的跨文化普适性上，去掉了中西方的调节作用。在假设和数据分析部分也进行了相应修订，请专家审阅。

7. 假设的编号：为了避免混淆，文化调节效应的两个假设，建议修改为假设 3 和假设 4。

作者回复：谢谢专家的宝贵意见，我们去掉了文化的调节效应，相应的假设也进行了修改。请专家再审阅。

8. 中介效应分析: **Bootstrap resampling** 次数为 5000, 建议至少 10000, 得出的结果更为稳健。

作者回复：非常感谢专家的宝贵意见，我们重新进行了数据分析，对文章相应部分也进行了修订，请专家再审阅。

9. 建议在所有研究中，报告完整的统计分析结果。比如研究一的回归分析，需要报告包括自变量、中介变量、以及控制变量的回归系数、SE, t 值, p 值等。

作者回复：非常感谢专家的宝贵意见，我们在文章中相应部分也进行了修改，请专家审阅。

10. 研究三表述不清晰：研究三招募西方被试，但是，在统计分析时，将中美样本数据进行比较。是否作者将研究三的美国样本数据与研究一的中国样本数据进行了比较？意义何在？需要补充说明。

作者回复：谢谢专家的宝贵意见，我们的表述的确不够清晰，因为想进一步明确文化在相关变量上是否本来就存在差异，所以在统计分析阶段是将西方数据和中国数据在几个核心变量上进行了比较，在文中也进行了说明，请专家审阅

第三轮

审稿专家 1:

感谢作者对上一轮评审意见的详细回复和充分考量。下面列出论文中依然存在的一些问题（按在正文中出现的顺序，而非按重要程度），尤其是在描述方法、汇报结果、基于统计结果做相应结论时，注意符合通行的规范。

1. 建议重写摘要，与现在的 package 和 Story 有很大冲突。例如，“已有大量研究考察人们对机器人做出道德决策产生算法厌恶”，但本文并没有考察机器人做出道德决策，而是在考察人类是否要对机器人做出道德考量及利他的行为选择。“个体的多元文化体验更广和更深”，但本文并未考察跨文化经历的这两个维度。

作者回复：谢谢专家的建议，已对摘要进行了修改，并标红，请专家审阅。

2. 引言中“于是，人工智能在技术和应用上的迅猛发展不可避免地带来了新的道德伦理问题。”这一句用“于是”有些突兀。是否可以写成：尽管人工智能在技术和应用上的迅猛发展大大提升了解决某些现实问题的效率与质量，但与此同时也不可避免地带来了新的道德伦理问题。（仅供参考）

作者回复：谢谢专家的建议，已按照您的建议修改，并在文中标红，请审阅。

3. 道德地位和利他行为的引入

作者在赋予机器人道德考量一节（1.1）比较轻描淡写地提到“利他行为是最基础的道德行为”。建议还是把这一小节留给机器人是否拥有道德地位。在小节 1.2 再提出，那么如何判断人类给予了机器人道德考量呢？有哪些研究指出，将对方视为道德对象的一个基本体现就是利他行为，并给出一些原因。接下来再说，哪些因素会影响人类对机器人的利他行为呢？一类因素是机器人的特点，一类是人类自身的特点（但研究较少）。

作者回复：谢谢专家的宝贵建议，已调整道德考量与利他行为的顺序，小节 1.1 专注说明机器人是否可能拥有道德地位，小节 1.2 则说明将机器人视为道德对象的一个基本体现就是利他行为，在正文中也对相关部分标红，请专家审阅。

4. 多元文化经历的引入

从道德地位和利他行为切到多元文化经历也有些突兀，虽然人类特征重要且研究少，但为什么要看多元文化这一个呢？是否可以说，既然利他行为是人类给予机器人道德考量的基本行为体现，哪些个体因素可以促进这种利他行为呢？利他的本质是接纳、包容、关爱、平等对待他人等（如有文献支撑），而且利他行为做为一种道德行为，其实是受很多非理性因素影响的。那么个体的 经历尤其是多元的开放的经历被发现可以增加包容...以及认知上的灵活性...因此，本文选择聚焦多元文化经历。然后再具体讲多元文化是什么，多元文化会影响什么，这些“什么”又会如何影响利他行为。

作者回复：谢谢专家的宝贵建议，已修改多元文化经历的引入，修改为“对机器人的利他行为是人类给予机器人道德考量的基本行为体现，哪些个体因素可以促进这种对机器人的利他行为呢？”，后续也按专家的建议进行了多元文化经历的引入，在文中已标红，请专家审阅。

5. p.26,“体验性(experience), 即拥有感知觉和情绪体验的感知能力”中的“感知觉”读不顺。

作者回复：谢谢专家的建议，已将本句修改为：

体验性 (experience)，即拥有感知情绪体验的能力，如感觉到饥饿、恐惧、痛苦、快乐和意识等。

6. p.28, “相比于西方国家的民众(如西方)” 似乎有问题。

作者回复：谢谢专家的建议，已将本句修改为：相比于西方国家的民众(如美国)。

7. “1.5 多元文化经历提升对机器人利他行为的跨文化普适性” 这一部分对文献的描述有些混乱。可以先提出问题，在不同文化中，这一关系链条有差异还是相似呢？再说有文献考察了相关问题。有些研究发现了文化差异，如... 还有些研究发现了没有文化差异，如... 因此，到底这一关系的方向和强度是否有文化差异性还是文化普适性，值得探讨。我们不提出具体假设，把它当个开放性问题。

作者回复：谢谢专家的建议，已修改这部分的文献综述，先提出问题，再说明前人研究中不一致的地方，并在文中标红，请专家审阅。

8. 在研究 1a 之前可以加一小段，说研究一的目的和 basic design，同时在中美两个文化中对三个假设加以检验。

作者回复：谢谢专家的建议，已在研究 1a 之前添加一段描述研究一的目的和实验设计，请专家审阅。

9. 描述量表时请保持一致：采用 6 点计分(1=“频率很低”，6=“频率很高”)；采用七点计分(从 1=“完全不同意”到 7=“非常同意”)

作者回复：谢谢专家的建议，已经全文修订，在量表描述时保持一致，请专家审阅。

10. “(4)人口学变量”这一段中“主体”与“个体”读起来有些 confusing.

作者回复：谢谢专家的建议，已修改为：现有研究对个体的自身特征对机器人利他行为的研究尚不充分，但已有研究指出个体特征可能会影响个体的道德边界。请专家审阅。

11. 表 1 和表 4 中对角线上的 1 应该全部移除，并在相应位置给用 scale 测量的变量提供 alpha，放在括号中。

作者回复：谢谢专家的建议，已经在表 1 和表 4 中增加了 α 一列。

12. 表 2、表 6 和表 8 请按规范格式重新整理（见附件的例子或任何一篇 decent journal 上的类似结果汇报）。注意：表 2 中，多元文化经历对心智知觉的影响，在纳入和未纳入控制时的数字完全一样，请纠正以避免这类基本错误。

作者回复：谢谢专家的建议，已修改，并再次修订了全文，避免此类错误。

13. 研究 1a 的结论“研究 1a 的结果表明多元文化经历可以正向预测对机器人利他行为，且心智知觉在其中起部分中介的作用，验证了本文的假设一”跟两个假设不匹配，请逐一说明哪个结果支持了哪个假设。

作者回复：谢谢专家的建议，已逐一说明哪个结果支持了哪个假设。

13. p.33, 表 5 这种简单的 t-test 有什么意义呢? 起码也得控制所有的人口差异等变量 (及多元文化经历), 再看中美被试在对机器人的心智感知与利他行为上有没有差异, 这样比较, 即便心智感知有差异, 也有可能是因为 sample differences, rating bias, etc. 导致的, 很难基于此下任何结论。因此 p.37 说“本研究也在一定程度上支持不同文化下的人们对机器人的心智知觉存在差异。”也是没有 convincing 根据的。

作者回复: 谢谢专家的宝贵建议, 只将中美样本做简单的 t-test 的确存在问题, 先已删掉这一部分的内容。

14. p. 34, “相对而言, 西方样本的中介模型的间接效应小于中国样本的间接效应”, 没有 statistical test 是不能做这种 Visually 比大小的。

作者回复: 谢谢专家的宝贵建议, 根据上一条建议, 已对这部分进行了删除。

15. “2.2.3 中西模型差异检验”应该加上“额外分析”additional analysis, 这样读者知道你检验的不是任何正式假设的一部分, 只是做了 1a, 1b 后不如再来比较一下。

作者回复: 谢谢专家的宝贵建议, 在中西模型差异检验前加上了“额外分析”, 请专家审阅。

16. p.38, “(SD = 4.49),。”多了一个逗号。“6.41(SD = 1.33)层”, 按前面的写法, 应该是 6.41 层 (SD)。“最后多元文化经历组被试 87 个、家乡文化经历组 83 个, 日常生活对照组 79 个”顿号和逗号混用。心智知觉的测量“均采用六点计分”没有说明 1 和 6 分别是什么。

作者回复: 谢谢专家的建议, 已经删掉了逗号, 将混用的逗号和顿号中的逗号替换成了顿号。

17. p.39, “、我觉得如果能加入机器人保护组织当一名志愿者是不错的选择”, 顿号应该在引号之外。对机器人利他行为的测量“采用六点计分”, 没有说明 1 和 6 分别是什么。“在告知被试此次问卷调查的目的以及数据信息的安全保障后, 被试填写知情同意书。”这两个分句的主语不一致, 比如应该是“在被告知此次问卷...后, 被试填写...”或“我们告知被试此次..., 然后请被试填写...”

作者回复: 谢谢专家的建议, 已修改顿号的位置, 并说明六点计分分别代表什么。最后将这部分修改为“我们告知被试此次问卷调查的目的以及数据信息的安全保障后, 然后请被试填写知情同意书。”

18. “3.1.4 研究程序”应该在“3.1.2 多元文化经历启动”和“3.1.3 测量工具”之前, 请按规范顺序汇报。

作者回复: 谢谢专家的建议, 已将“3.1.4 研究程序”放在“3.1.2 多元文化经历启动”和“3.1.3 测量工具”之前, 请专家审阅。

19. p.39, 在利他和心智两个结果上组间两两差异皆不显著, 就不能说“高于”, 再说事后检验不显著, 这种 visually identified difference 是不可以汇报的, 看上去两个数差再大, 统计检验不显著, 就只能说不显著。

作者回复: 谢谢专家的建议, 已经去掉了这部分的表述, 请专家审阅。

20. pp. 39-40, 研究二引言中说体验性感知比能动性感知更容易引发利他行为。而在结果中发现多元文化经历对能动性的影响在 $p = .10$ level 上显著; 讨论中又说“这在一定程度上表明相对于体验性, 多元文化经历丰富的人们更可能感知到机器人拥有主体能动性”。

作者回复: 谢谢专家的宝贵建议, 已修订主体性和能动性这一部分的表述, 请专家审阅。

21. p.40 总讨论中“结果发现, 对机器人的心智知觉和利他行为存在一定的文化差异: 西方文化下的被试对机器人的心智知觉和利他行为显著高于中国文化下的被试”这个结论是没有根据的。首先, 在利他行为上, t -test 是未发现中美差异的; 其次, 如前所述, 即便对于心智感知, 这种 simple t -test 也不能得出可靠的结论。

作者回复: 谢谢专家的宝贵建议, 我们去掉了利他行为的 t -test, 在总讨论中也删掉了相应的内容, 请专家审阅。

22. p.41, “通过元分析的技术验证是否该效应是否稳健”, 两个是否。

作者回复: 谢谢专家的宝贵建议, 已修改为“通过元分析的技术验证该效应是否稳健”。

23. p. 42, “5 5 结论”两个 5。请仔细检查全文, 避免语法、表达、规范等各方面的错误。

作者回复: 谢谢专家的宝贵建议, 已修改, 并通篇订正, 修改了语法、表达、规范等各方面的错误, 请专家审阅。

.....

审稿专家 3:

作者对审稿意见作了回应与修改, 解决了相应的问题。但文章写作方面仍需润色和斟酌。

1. “社会心理学如何参与人工智能革命浪潮”, 社会心理学作为心理学的一个子学科, 为何要参与人工智能革命浪潮? 作者想表达的是什么?

作者回复: 谢谢专家的宝贵意见。作者本意是想说明社会心理学对于如何更好地进行人机交互具有重要作用, 但可能表述有些模糊, 已进行了修改, 请专家审阅。

2. “被试会更快地响应机器人的求助请求”, 介绍他人的已有研究结果时, 不要使用“被试”, 需要使用该研究样本所代表的群体词汇

作者回复: 谢谢专家的意见, 已修改。

3. “关注人类自身特征对机器人利他行为的研究相对较少, 需要进一步的理论建构和实证探索”, 建议减少这种空洞的描述, 需要具体说明哪方面的理论建构和实证探索

作者回复: 谢谢专家的意见, 已具体说明。

4. 多元文化经历中的文化背景, 是否影响个体的利他行为? 是否某些文化的经历可能比另一些文化的经历更促进利他行为? 尽管这个研究未必需要回答这一问题, 但作者可以进行理论上探讨。另外, 为便于未来研究重复本研究结果, 建议作者报告留学生的多元经历包括哪

些文化或国家（比如百分之多少被试在美国留学）。此外，数据收集时，被试在留学还是已经回国？

作者回复：谢谢专家的宝贵意见，对多元文化经历中的文化背景是否影响个体利他行为这一问题在正文中的表述是作为一个开放性的问题进行探索。此外，当时并未收集留学生的多元经历包括哪些文化或国家以及是否回国等背景信息，因此未在文中具体说明。

5. 研究 1a 招募海外生活经历 3 个月以上的被试，为什么研究 1b 调整为海外生活经历 6 个月以上的被试？时限的标准是什么？需要引用文献支持

作者回复：谢谢专家的意见，已修改，加入了参考文献。

6. 研究二需要说明是否为中国被试

作者回复：谢谢专家的意见，已说明。

7. 研究二仅部分验证假设，但作者的表述较为含糊，建议作者在结果和讨论部分清晰写明。一方面，有助于读者了解哪些因果关系得到了验证，而哪些没有；另一方面，有助于未来研究进一步探讨多元文化经历与利他行为的因果关系，比如是否这一关系受某种因素调节，仅在某种条件下这一因果关系才成立等。

作者回复：谢谢专家的宝贵意见，在研究二的结果和讨论部分进行了更详细的说明。

编委复审

编委意见：我觉得可以了。

作者回复：十分感谢您对拙作修改的认可和审稿专家们提出的若干建设性意见。

主编终审

主编意见：建议作者考虑：将文章题目“多元文化经历提升人类对机器人的利他行为：心智知觉的中介作用”改为“多元文化经历提升人类对机器人的利他行为及心智知觉的中介作用”是否更合适？

作者回复：谢谢主编专家的意见，已将题目“多元文化经历提升人类对机器人的利他行为：心智知觉的中介作用”改为“多元文化经历提升人类对机器人的利他行为及心智知觉的中介作用”。