《心理学报》审稿意见与作者回应

题目：中西方文化差异对虚拟人道德责任判断的影响
作者：闫霄，莫田甜，周欣悦

第一轮

审稿人 1 意见：

意见 1.1：作者似乎没能准确把握 4 个实验的展开逻辑：本研究的核心主效应究竟是文化差异还是虚拟人-自然人之间的差异？如果核心是虚拟人-自然人之间的差异，那为什么正文汇报的简单效应都是跨文化差异？
回应：非常感谢您的问题。我们研究的首要核心是这**两个变量的交互效应而不是单纯的主效应。我们关注的重点是在虚拟人组中文化差异的简单效应。**本研究的主要贡献在于引入了文化差异，主要关注文化差异对于虚拟人道德惩罚的影响，自然人只是我们引入的一个跟虚拟人的参考对象。这也就是为什么我们汇报的简单效应更加聚焦于文化差异的影响。

意见 1.2：如果核心是文化差异，那为什么引言部分全部是在论证虚拟人-自然人之间的差异？
回应：感谢您指出这个问题。根据您的指导，我们在引言部分已补充文化差异的内容，详见正文第二段、“1.4 心智感知的文化差异”以及“1.5 道德责任判断的文化差异及感知心智能力的中介作用”。

意见 1.3：论文的“东西方文化差异和虚拟人道德责任判断”这一主标题颇令人费解，“和”字前后的东西是什么关系？本研究要考察的是两个平行的科学问题吗？
回应：感谢您的建议，题目已修改为“中西方文化差异对虚拟人道德责任判断的影响”。

意见 1.4：审稿人没能在本文的引言部分获取有关“虚拟人-自然人道德判断差异”的前人研究基础，按照常规的研究推进思路而言，这应当是本研究首先需要检验和讨论的效应，然后才有根据进一步推进到跨文化差异的探索。

回应：感谢您的建设性意见。根据意见 1.1，本文首先已修改为以文化差异为核心主效应进行论证，并且在“1.2 对虚拟实体的道德责任判断”部分中补充了对虚拟实体道德责任判断的研究基础，并根据心智感知理论进一步推导到文化差异的探索。

意见 1.5：实验 2 单独考察了感知心智能力的文化差异，这里存在两个重要问题：①感知心智能力何以作为一个重要因变量、单独设置一个实验进行考察？②感知心智能力的文化差异为何存在？何以仅在虚拟人身上存在而在自然人身上不存在？

回应：感谢您的问题。①结合您的建议，我们删除了原文章中的实验 2(感知心智能力的文化差异)，将其作为现文章中实验 3(感知心智能力的中介效应)的预实验，详细内容放到了“附录二”中。感知心智能力本身也属于一个比较重要的变量，在 Gray 等人(2007)发表在 Science 上的论文中，作者就提到了感知心智能力的重要性。而且，感知心智能力属于本研究中一个

重要的中介变量，提供了解释机制，并且目前没有相关研究考察对虚拟人心智能力评价的文化差异。②之前的研究表明，**相比于西方人，东方人存在更高的拟人化倾向(Letheren et al., 2016)和换位思考能力(Dietze & Knowles, 2021)等**。因此，东方人会认为虚拟人和自然人更相似，从而对虚拟人的心智能力评价更高；而对于自然人来说，因为东西方人都会认为自然人具备很高的心智能力，因此他们之间并不存在差异。这部分的详细解释放在“1.4 心智感知的文化差异”部分中。

意见 1.6：《心理学报》今年早些时候在线录用和预出版的《算法歧视比人类歧视引起更少道德惩罚欲》一文（现已正式出版，2022 年第 9 期）和本文考察的内容是很类似的，不知作者在投稿前是否关注到此文？

回应：感谢您提示我们这篇研究。由于《算法歧视比人类歧视引起更少道德惩罚欲》(许丽颖 等, 2022)一文在本文投稿期间发布，在投稿前并未参考。根据您的提示，我们也在引言部分引用了这篇研究。许丽颖等(2022)表明，相对于人类歧视，人们对算法歧视的道德惩罚欲更少。潜在机制是人们认为算法(与人类相比)更缺乏自由意志，且个体拟人化倾向越强或者算法越拟人化，人们对算法的道德惩罚欲越强。

（1）首先，这篇文章的结论能够一定程度上支持我们的研究发现，因为该研究用中国的样本数据发现高拟人化程度的算法会增加人们对其的道德惩罚欲。如果我们把虚拟人看成拟人化程度比较高的算法的话，的确在中国文化中，人们更愿意惩罚犯错的虚拟人。因此我们的结论跟这篇文章的结论在这一点上是一致的。

（2）我们的研究跟这篇文章的主要区别有三个：①我们的主要关注点在于文化差异对虚拟人道德惩罚的差异，因此我们的样本有东方人也有西方人。②在中介机制上，本文考察感知心智能力的文化差异，与许丽颖等(2022)一文考察的自由意志信念不同。感知心智能力还包括了个体的自我控制、思考在内的自主力以及情绪感受在内的感知力(Gray et al., 2007)，因此感知心智能力比自由意志更加底层而且包括的范围也更加广泛。③虚拟人虽然可以被看作是高拟人化的算法，但是实际生活中社交媒体上面现有的虚拟人的各种动态和活动都是由背后的人或者公司来进行的。从这一点来看，虚拟人其实只是一种代理。从这一点上看，虚拟人和算法是存在差异的(Arsenyan & Mirowska, 2021; Moustakas et al., 2020; Sung et al., 2022)。

意见 2：引言中对文化差异的论证极为单薄且乏力。“1.1 虚拟实体的道德责任及文化差异”部分似未说到文化差异，怎么最后一句就忽然引入“跨文化差异”了？从这一部分来看，本研究要考察的不应该是虚拟人-自然人的差异吗？引言中涉及文化差异的论述仅在“1.3 文化差异、虚拟人的心智感知和道德责任”中随意堆叠式地、非常简单地提了 Willard 和 McNamara(2019)、Haslam 等(2008)、Dang 和 Liu (2022)、Wu 和 Keysar (2007)以及 Letheren et al. (2016)共 5 篇参考文献，其中只有 2 篇和“机器人心智感知&文化差异”有关，却还没能细致告诉读者“中国人对机器人五种感官水平的评价相对更高”“中国人对机器人社会性的感知更高”如何体现、意味着什么？“研究者基于内隐理论的解释”也应作为文化差异的一个重要的理论基础进行阐述，而不是一笔带过。在说了半天虚拟人之后，假设中却没有关注虚拟人-自然人之间的差异，而是呈现了文化（中-西）×道德主体（虚拟人-自然人）交互中的文化差异方面的简单效应。如果说假设 1 的前半部分勉强可以接受的话，假设 1 的后半部分何以提出？总之，如果作者要考察文化差异，绝不是这样单薄的介绍、加上一句“本文认为”就可以推出本文的假设的。

回应：感谢您指出的这些问题，我们之前引用的文献确实无法详细论证文化差异。根据您的

建议，我们已从文化差异的角度重新推导理论框架。详细内容已放在“1.4 心智感知的文化差异”以及“1.5 道德责任判断的文化差异及感知心智能力的中介作用”部分中。我们增加了参考文献，也增加了逻辑推理的部分。

意见 3：摘要。铺垫过长，对实验的概述过于松散、没能综合而系统地呈现本研究的发现和推进之处，缺少研究意义。

回应：非常感谢您的建议。根据您的意见，我们已按照《心理学报》的摘要格式修改，详见摘要部分。

意见 4. 1：虚拟人和人工智能机器人有什么区别？特殊性在哪里？依托社交媒体平台也算是虚拟人与人工智能机器人的不同？人工智能机器人和真人形象不相似？即使算是不同，那么作者指出的这两点（至少是非本质的）“特殊性”何以至于对道德责任判断产生影响？

回应：感谢您的问题。本文已补充聚焦社交媒体上的虚拟人而非人工智能机器人的原因，详见“1.1 虚拟人的定义及研究范畴”部分。目前的研究者认为，虚拟人和人工智能机器人存在一些相似之处，但是又有很多本质的区别。

■ 它们的相似之处在于：

虚拟人和人工智能机器人在创造和运营过程中都采用了人工智能技术(Sung et al., 2022; Kim & Jo, 2022)，有形体的人工智能机器人和虚拟人都可以模仿自然人的行为、情绪和交互能力(Sullivan & Wamba, 2022)。

■ 它们的不同之处在于：

(1)从本质上，已有文献将虚拟人定义为计算机生成的拟人化图像(Arsenyan & Mirowska, 2021; Moustakas et al., 2020; Sung et al., 2022; Kim & Jo, 2022)，表明其具有类人的外观；而通过人工智能创造的机器人不一定具有类人的外观；又因为视觉外观是感知心智水平的重要线索(Krumhuber et al., 2015)，因此我们认为人们对虚拟人和人工智能机器人的心智感知会存在差异，从而影响道德责任的判断。

(2)从功能上，虚拟人已被作为一项媒体技术广泛应用于日常生活中(Kim & Jo, 2022)，它们可以用语言、姿势，甚至表情来和人们进行面对面的交流互动(Volante et al., 2016; Kim & Jo, 2022)；但人工智能机器人基于算法系统(Russell & Norvig, 2002; Sung et al., 2021)，大部分人工智能机器人主要用于替代人类解决问题，因此在心智能力的情绪感知维度上可能与虚拟人存在差异，如果人工智能机器人被描述为一个情感实体，那么也可能被赋予类人的心智(Lee et al., 2020)。

(3)从应用场景上，目前虚拟人主要是作为新的营销和宣传工具出现在媒体平台(Arsenyan & Mirowska, 2021; Moustakas et al., 2020)，而人工智能机器人的应用场景更加广泛，不同类型的机器人存在不同的功能和应用场景，人们对其感知取决于不同场景，因此我们将研究对象聚焦社交媒体上的虚拟人而非人工智能机器人。

(4)从行为上，随着越来越多的虚拟人出现在人们生活中，道德伦理问题也逐渐浮现，例如人们会看到虚拟人 Bermuda 盗用他人账号、“四禧丸子”抄袭、虚拟偶像团体 A-SOUL 侵权等等报导，而很少听说人工智能机器人进行类似的不道德行为，因此人们在对其道德责任的判断上可能也会存在差异。

(5)从关系上，虚拟人与人工智能机器人存在技术上的交叉，虚拟人是利用人工智能技术创造的一种新的代理形式(Sung et al., 2014)，人工智能技术可以赋予其更真实的外表、更自然的表达(Loveys et al., 2020)，但其不限于人工智能技术，还包括计算机图像生成技术、动捕技术等。

综上所述，虚拟人与人工智能机器人在本质、功能、应用场景和行为上都存在差异。虚拟人是人工智能技术和其他科学技术的结合体，能够通过媒体技术和人类进行交互，并且存在不道德行为的现实案例，而其拟人化外观和社交属性都会成为其区别于人工智能机器人的特点，来影响人们的道德责任判断。

意见 4.2：如果要考察的是对虚拟人-自然人之间的道德责任判断差异，那么引入心智感知还算合理；但如果要考察的是文化差异，为什么要考虑心智感知作为文化差异的心理机制，是不是过于牵强了？可以说揭示的仍是现象，没有深入到机制。

回应：感谢您的问题。已有的研究表明，文化差异会影响到人们的拟人化倾向(Letheren et al., 2016)，并且东方人比西方人更擅长换位思考(Wu & Keysar, 2007)，所以东方人更倾向于拟人化的思维方式。也就是相比于西方文化来说，东方文化会让人更倾向于把物体赋予上人的特质和人的独特能力。因为人们倾向于相信心智能力是人类(或者近似人类)的一种特有的能力(Haslam, 2006; Haslam & Loughnan, 2014)，那么也就意味着东方文化会导致人们更加倾向于认为物体也具备心智能力。从这一点来推论，相比西方人来说，东方人会认为虚拟人更加类似自然人，具备自然人特有的心智能力。接下来，因为心智能力是道德判断的本质，与道德判断息息相关(Gray et al., 2012; Ward et al., 2013)，我们认为，东方人给虚拟人赋予更多的心智能力，从而也会影响对虚拟人的道德判断。

意见 4.3：此外，研究逻辑链中的“心智感知→道德判断”这一路径的证据缺失，“1.2 虚拟实体的心智感知与道德责任”部分第一段中的两个“因此”是毫无逻辑的。

回应：感谢您的问题和建议。根据您的指导，我们补充了这部分的证据。感知到的心智能力会影响道德判断，这也就是为什么我们对于心智能力缺乏的人，比如未成年人或者某些精神病患者的犯罪的惩罚比较宽容(Gray et al,, 2012)。大量的研究证据也支持了这个论断。心智感知主要分为两个维度，一个维度是感知力，也就是个体是否感受到快乐、痛苦、愤怒等(Gray et al., 2007)。对别人痛苦等情绪的感受被认为是道德决策的关键(Greene et al., 2001; Haidt et al., 1993; Hume, 1751)，如果个体能感受那么意味着个体可以对别人产生共情，因此不太愿意把痛苦施加给别人。另一个维度是自主力，也就是个体是否能够行动、计划、实施自我控制、记忆、交流和思考，反映了个体的行为背后是否有意图和目的。如果我们觉得这个个体是有意犯罪，跟我们觉得个体是无意犯罪，对他们的判决是完全不一样的。这一部分的相关理论已进行补充，并放在了“1.3 心智感知理论”及“1.5 道德责任判断的文化差异及感知心智能力的中介作用”部分。

意见 4.4：小问题：①引言第 1 段最后的问题切入非常突兀，应结合现状及同领域研究简要论述；②第 2 段的第一句话和第 1 段的内容重复了；③第 2 段“虚拟人 Bermuda 非法盗用了他人的账号并且擅自删除账号内容”真的是虚拟人干的吗？从引用的链接来看似乎存疑；而且对于类似的新兴事物，应细致介绍一件典型事件，不能全部一笔带过；④第 2 段为何忽然引入文化差异？应有理论或现实证据表明可能存在潜在的文化差异吧？⑤引言第 3 段似无必要；⑥第 4 段“但许多学者认为，道德责任不能够被解释为一个单一概念，而它涉及相对多元的定义”一句中什么叫“单一概念”？哪些学者对亚里士多德产生了质疑？⑦第 5 段“学术界在虚拟实体是否能够承担道德责任的问题上存在广泛争议”形成了两种“思想流派”吗？这方面的哲学伦理学文献还是十分丰富的，需要做更细致的阐述；⑧第 10 段说“虚拟人大多依托社交媒体平台与人类进行交互，背后可以是人工智能算法、独立个体或公司集团等”，那么虚拟人背后有可能是真实的人或团体？⑨第 11 段“由于心智能

力的感知是道德判断的本质(Gray et al., 2012)”，这句话太宏大了，应做更深入的阐释，而不是简单把标题翻译过来。

回应：非常感谢您的建议性意见。①②部分已在文中修改；③在报道中，虚拟人 Bermuda 确实非法盗用了另外一个账号并且擅自删除账号内容，从理性的判断，这背后肯定是真实人类做的而虚拟人只是代理。如果每个人都是这样理性的判断，那么对于虚拟人的惩罚应该跟对于真实人类的惩罚是一样的。我们的研究关注的是人们对虚拟人的这种非理性的思考会影响到我们对虚拟人的道德惩罚；④文化差异的引入部分已重新补充完善，详见“1.4 心智感知的文化差异”；⑤已删减；⑥⑦这里的表述不够清晰，已重新修改；⑧我们同意您的意见，虚拟人的背后确实有可能是真实人类或团体，比如许多社交媒体虚拟人发布的视频甚至直播都是通过真人进行实时动捕的。为了测试被试在对虚拟人幕后的判断是否会影响我们研究的主要结果，我们新增加了一个实验 4 操纵了虚拟人背后的主体类型。结果发现，无论虚拟人背后的主体是真实人类还是人工智能，中国人都会比西方人给出更高的道德责任评价；⑨已按您的建议修改表述。

意见 5. 1：各实验各组采用 100 名被试，建议作者增加汇报被试量确定的原由，计算并汇报这一样本量下的统计检验力。

回应：非常感谢您关于样本量的建设性意见。根据您的建议，我们已在“实验 1”中的“2.1 实验设计与样本”第一段中对样本量的确定标准给出说明：“实验采用 G*Power 3.1 软件(Faul et al., 2007)计算实验所需样本量。对于本实验适用的双因素方差分析，当显著性水平 α 为 0.05 且效应量为中等效应时(f = 0.2 到 0.25 之间)，要达到 95%统计检验力所需要的样本量在 279 到 434 之间。为保证实验具有足够的样本量且为统一标准，本研究的所有实验按照每个条件下 100 名被试的标准招募，且采用原始数据进行分析。但由于线上平台存在同时作答以及未完成作答等情况，会出现多于或少于既定被试数量 1-2 个的情况。”

意见 5. 2：实验材料中，真人组-虚拟人组的社交网站头像为什么不设置成一样的？这会不会成为混淆变量？

回应：感谢您的问题。实验材料中社交网络头像的真人和虚拟人是按照同一个人的照片为模板用软件进行制作的，这样保证了面部表情的一致。虚拟人的头像也基本符合互联网上的虚拟人的常见形象设置。为了保证实验材料的有效性，我们做了前测，招募了 195 人随机分成两组对两种头像的吸引力、可信度、专业度、熟悉度以及态度。结果表明，头像的吸引力($M_{虚拟人}$ = 4.93, SD = 1.39 vs. $M_{真人}$ = 5.22, SD = 1.12; t (193) = 1.60, p = 0.112)、可信度($M_{虚拟人}$ = 4.41, SD = 1.25 vs. $M_{真人}$ = 4.32, SD = 1.03; t (193) = −0.54, p = 0.589)、专业度($M_{虚拟人}$ = 4.43, SD = 1.40 vs. $M_{真人}$ = 4.25, SD = 1.50; t (193) = −0.87, p = 0.384)、熟悉度($M_{虚拟人}$ = 1.13, SD = 0.45 vs. $M_{真人}$ = 1.26, SD = 0.71; t (193) = 1.46, p = 0.146)和态度($M_{虚拟人}$ = 4.29, SD = 0.86 vs. $M_{真人}$ = 4.38, SD = 1.00; t (193) = 0.64, p = 0.523)都没有显著差异。**值得提到的是，我们的核心假设并不是虚拟人和真人的主效应，而是文化差异的调节效应，简单效应显著的虚拟人组的中国被试和西方被试看到的是同样的图片和文字介绍。**

意见 5. 3：因变量指标中翻译过来的“Rico 应该承担多少指责”似乎不太符合中文表述习惯，两个题目的相关才 0.76 有点奇怪，请在两种文化下分别报告两个题项（承担多少责任？承担多少指责？）之间的相关系数。

回应：感谢您指出这个问题。首先，如果去掉这个题项，我们的主要结果还是保持不变。其

次，在相关系数上面，按照您的建议，我们在两种文化下分别报告了两题的相关。可以看出，在中文中两者的相关并不会比英文中两者的相关要显著的更低。

对于修改前的实验 1，两个题项的整体相关系数为 0.76，$p < 0.001$；在中国被试中，相关系数为 0.66，$p < 0.001$，在西方被试中，相关系数为 0.81，$p < 0.001$；

对于修改前的实验 3，两个题项的整体相关系数为 0.84，$p < 0.001$；在中国被试中，相关系数为 0.72，$p < 0.001$，在西方被试中，相关系数为 0.88，$p < 0.001$；

对于修改前的实验 4，两个题项的整体相关系数为 0.71，$p < 0.001$；在中国被试中，相关系数为 0.81，$p < 0.001$，在西方被试中，相关系数为 0.66，$p < 0.001$。

另外，鉴于您提到之前实验的测量条目存在表述上的问题，**我们重新补充了实验 2，采用新的道德责任量表(Cameron et al., 2010; Cronbach's α = 0.81)来测量，复制出了实验 1 的结果**，说明我们得到的效应并不是因为测量条目的表述问题引起的。此外，在更新版的实验 4 中，我们根据第二位审稿人的意见将这个条目的表述更改为“多大程度上应该归咎于 Rico？”。两个题项的整体相关系数为 0.71，$p < 0.001$；在中国被试中，相关系数为 0.67，$p < 0.001$，在西方被试中，相关系数为 0.68，$p < 0.001$。相关系数依然不算很高，但两种语言表述下的相关性不存在显著差异。

意见 5. 4：心理学报要求报告差异的置信区间，至少在简单效应分析中应报告相应的文化差异的置信区间。

回应：感谢您的建议。我们已在正文数据分析的结果部分补充了差异的置信区间。

意见 5. 5：多个实验中存在类似“更重要的是，①文化类型和博主类型的交互作用能够显著影响道德责任的判断”的表述，请问“更重要”何出此言？②“交互作用显著影响 XXX”又是什么意思？③条形图的图注也宜再考量，图中能看出来交互显著吗？

回应：非常感谢您的修改意见。①使用“更重要的是”这样的表达是为了强调我们应该更加关注交互效应而不是主效应。根据您的建议，我们去掉了这个短语；②“交互作用显著影响道德责任判断”也按照您的建议进行修改，具体描述为“文化类型和博主类型的交互作用对道德责任判断具有显著的影响”；③感谢建议，已在图中补充图注。

意见 5. 6：实验 2 结果中“文化主效应边际显著”是什么？

回应：感谢您的问题。“文化类型的主效应边际显著”指的是对于整体样本来讲(包括真人组和虚拟人组)，两种文化的被试对 Rico 的心智能力评价存在边际显著的差异($p = 0.049$)。具体而言，西方被试比中国被试对 Rico 心智能力的评价更高。当然这种主效应并不是我们关注的重点，我们的研究假设更关注交互效应和简单效应的结果。在本轮修改中，我们删除了原文章中的实验 2，将其作为现文章中实验 3 的预实验，详细内容放到了“附录二”中。

意见 5. 7：实验 3 中的“虚拟人的熟悉度”和“事件的严重性”的文化差异都显著，①是否作为协变量纳入中介模型中进行控制？②有调节的中介不成立，这两个潜在混淆变量在道德责任判断文化差异中的简单中介效应是否成立？③“排除替代中介解释”是排除对谁（文化差异 vs. 自然人-虚拟人差异）的中介解释？还是说关注点又变成“文化×道德主体”的交互了，那么实验 4 怎么又不考虑交互了？

回应：感谢您指出这些问题。①将“虚拟人的熟悉度”、“事件的严重性”、性别和年龄纳入控制变量后，并不会改变有调节的中介效应的数据结果(indirect effect $\beta = 0.35$, $SE = 0.11$,

95%CI = [0.16, 0.58])；②从数据结果上看，这两个混淆变量的简单中介效应均不成立，在虚拟组，事件严重性的简单中介效应不成立(indirect effect β = 0.06, *SE* = 0.14, 95%CI = [−0.20, 0.36])，熟悉度的简单中介效应也不成立(indirect effect β = 0.19, *SE* = 0.18, 95%CI = [−0.18, 0.55])；即使简单中介效应成立，有调节的中介不成立也不足以作为解释机制，因此正文中没有继续汇报简单中介效应；③因为结果表明，真人组不存在文化差异，因此这里主要排除的是对于虚拟人组文化差异的替代中介解释。另外，因为前面的四个实验都表明真人组没有文化差异，因此更新后的实验 5 我们进行了简化，聚焦虚拟人组进行研究。

意见 5. 8：实验 4 中引入了道德惩罚作为因变量，其中一个指标是“是否应该被罚款”，罚谁的款？虚拟人的款？怎么罚？

回应：感谢您的问题。这里表达的是对账号运营主体进行罚款，不管是用虚拟人还是真实人的照片作为社交网络的账号头像，背后都有一个利益主体，可能是真人个体，也可能是公司或者工作室，因此对虚拟人进行罚款是可能并且可行的。类似的惩罚(封号、取消关注)和罚款具有一致的结果。当然，可能每个被试对于虚拟人账号背后的运营方式的理解不一样，就这一点，我们也增加了一个新的研究(实验 4)来考察这种理解是否存在影响。结果发现，无论虚拟人背后的主体是真实人类还是人工智能，中国人都会比西方人给出更高的道德责任评价。因此，我们的主要结果并不是由于这种不同的理解来引发的。

意见 5. 9：所有实验都是在线实验，如何保证被试作答质量？是否设置了测谎题（注意检测题）？

回应：感谢您的问题。问卷设置了注意力检测题项，但由于按照注意力题项筛除被试后结果保持不变，为了按照《心理学报》要求保证数据的完备性，因此统一采用原始数据进行分析汇报。

意见 5. 10：受教育程度的“博士及以上”中的“以上”是什么？

回应：感谢您的问题。理论上博士研究生是最高学历，我们考虑到，有些中国人可能会把博士后误解成为是一个学历，因此我们这里保留了“以上”。否则可能会有一些存在这种误解的被试感觉自己的学历没有被包括。

意见 6. 1：结论应放在正文最后呈现。

回应：按照您的建议，我们进行了修改。我们参考了《心理学报》最近发表的论文，发现有一种写法将结论放在了总讨论部分中的第一段(如王丽丽和董梦璐, 2022; 冯文婷 等, 2022; 陈增祥 等, 2022)，一种是将结论另做一部分放在正文最后(如许丽颖 等, 2022; 陈茗静 等, 2022; 孙博, 2022 等)。目前，我们将结论放在了正文最后呈现，详细内容见“8 结论”。

意见 6. 2：目前的讨论部分对于本研究发现的讨论几乎全是空话，未见明确的理论贡献，如：“因此，基于前人的研究，本文将研究对象从人工智能拓展到虚拟人，并研究东西方人对虚拟人道德责任判断的差异，丰富了道德判断与文化差异的相关文献。”又如“本文还以此作为中介机制，推演到对其道德责任承担判断文化差异，进一步丰富了心智感知理论和道德责任理论在虚拟实体上的相关研究。”

回应：感谢您提出的问题。我们已重新修改理论贡献部分，详见“7.1 对虚拟人的心智感知”

及“7.2 对虚拟人的道德责任判断”部分。

意见 6.3：讨论部分还需结合上述引言的修改建议和意见进行调整，围绕研究的核心关注点展开。

回应：感谢您的建议。我们已重新修改讨论部分内容，详见“7.1 对虚拟人的心智感知”及“7.2 对虚拟人的道德责任判断”部分。

……

审稿人 2 意见：

意见 1：在问题提出部分，应该进一步总结和提炼相关研究的现状和局限总结，说明选题的重要性并指出论文的主要贡献理论。例如，与人工智能和机器人有什么不同？对这种/些不同的探讨有哪些理论意义和实践价值。P.6 第二段的内容（随着越来越多的虚拟实体进入我们的日常生活.....）可以考虑提前。

回应：感谢您的建议。我们同意您的看法，并且按照您的提示将引言部分进行修改，详见文章前两段及“1.1 虚拟人的定义及研究范畴”。

意见 2：1.1 的题目是“虚拟实体的道德责任及文化差异”，但是该段中并没有涉及文化差异的内容，需要做出相应修改使段落内容和题目一致。

回应：感谢您的建议。我们大量的增加了文化差异的部分，文化差异的内容详见“1.4 心智感知的文化差异”以及“1.5 道德责任判断的文化差异及感知心智能力的中介作用”部分。

意见 3：1.1 中应先对道德责任的基础文献和理论进行综述和梳理，目前这部分内容存在欠缺，前文虽有涉及，但都较为零碎，系统性和归纳不足。

回应：根据您的建议我们进行了修改，道德责任的相关文献已在“1.2 对虚拟实体的道德责任判断”里进行综述和梳理。

意见 4：1.1 中提到了对虚拟实体是否能够承担道德责任的两种不同观点，对于第二种观点的交代浅尝辄止，需要说明这一学派的理论基础是什么，从而可以更清晰的阐释出产生不同观点的根源在哪里。这里引用的一个主要文献 Stahl, 2006，其论文主要围绕的是 computer，而不是人工智能或机器人，文献引用需要更严谨一些。

回应：我们同意您的看法，这部分的表述确实不够清楚，已在“1.2 对虚拟实体的道德责任判断”中调整修改，对于 Stahl, 2006 引文已不再提及。

意见 5：1.2 第一段最后一句话“因此，对一个实体心智能力的感知能够判断其是否应该承担道德责任。”①这个观点作为后面假设提出的基础，其合理性和充分性不足，需要进一步论证。心智能力的感知是道德责任承担的充分条件吗？②类比人类的情境，是否存在有充足心智能力，但是由于其他原因（如 intentionality, locus of control）等导致不该承担道德责任的情况？

回应：感谢您的问题。①根据您的建议，我们更加详细论述了心智能力为什么是道德责任承

担的基础，这个看法跟很多已有研究的观点是一致的(Gray et al., 2007; Waytz et al., 2010; Dietze & Knowles, 2021)，这部分内容在“1.3 心智感知理论”中详细陈述。②心智能力本身就包含了意图、计划和发挥自我控制的能力的这些维度(Gray et al., 2011)。

意见 6：1.3“这些研究表明，不同文化背景下，人们对其他个体在类人特征的看法上并不一致”该语句存在歧义，这里“其他个体”是指什么含义？

回应：感谢您指出这个歧义。这里表述不够清楚，这里的“其他个体”指的是“非人类的实体”，该部分的内容已经重新修改。

意见 7：论文 2 个假设的提出都比较仓促，条理不清，论证不足。例如，作者列举了一些研究发现，然后就提出“基于这些研究发现进行推断认为，东方人可能对虚拟人物的心智能力评价更高”这样的说法，无法令人信服。作者前面提到研究认为心智能力包括很多维度（如，计划、自我控制和感受情绪等），东西方消费者对虚拟人心智能力评价的差异主要体现在哪些维度上，这些维度是主导道德责任判断的因素吗？在假设提出中，需要围绕以上问题做出更细致的分析和阐释。

回应：感谢您指出的问题，在假设推导部分已重新修改完善。以往研究主要将**心智能力划分为自主力和感知力两个维度**(Gray et al., 2007)。在修改后的实验里，我们分别从这两个维度进行了分析，并且发现这两个维度都可以单独作为中介。但是平行中介分析的结果显示，感知力比自主力有更强的解释力度，这与近期的相关研究也是吻合的(Dietze & Knowles, 2021)。

意见 8：由于论述和实验中都是基于中国消费者，建议在假设提出时，采用“中西方”，而不用“东西方”。

回应：感谢您的建议，已作修改。

意见 9：假设 1 中的后半段“而当看到真人进行同样的不道德行为后，东方人和西方人在道德责任判断上没有文化差异”，这一观点在论述中并没有进行说明和阐释，却直接出现在了假设中。没有什么其他因素（东西方差异）会导致真人时的道德责任判断也是存在差异的吗？比如内外部归因、思维方式等。因此，假设 1 的陈述方式不严谨，需要修改，并做出必要的补充说明。

回应：我们同意您的意见，因为真人在道德判断上的文化差异不是本文的研究重点，我们将假设 1 的后半段删掉。另外，在我们的五个研究中没有发现中西方人在真人的道德责任判断上出现文化差异。这可能是因为这些道德问题主要出现在社交媒体或者互联网之上，而且涉及到的道德行为没有太多的文化特异性。我们猜想，如果替换别的情境或者道德行为可能会出现文化差异，但是这也不是本文的研究重点。因此我们也在讨论中增加了对这部分的讨论，详见“7.3 研究局限与未来研究方向”中的第四段。

意见 10：在多个实验中，女性被试占多数且中美实验组性别比例不平衡（如，实验 1 中国组被试女性 75%，实验 2 英国被试女性超过 80%），年龄也存在差异，需要说明这种被试构成是否会影响结果。虽然作者提到将性别加入后并不会影响实验结果，仍需要报告控制年龄和性别后的具体检验结果。

回应：感谢您的建议。已在文中补充加入控制变量后的结果，详细内容见“附录三”。结果表明，加入性别和年龄等控制变量后并不影响原先结果的显著性和方向。

意见 11：实验中分别采用基于 twitter 和基于微博的虚拟人物为考察对象。两国消费者对两种不同平台上对虚拟人物的运作机制和管控等背景信息的理解是否会影响被试对虚拟人为什么会出现不道德的行为以及如何对道德责任进行推断？这是否会对实验结果产生影响，需要做出必要的解释和说明。

回应：感谢您的问题。由于本文研究的核心问题在于对社交媒体虚拟人道德责任判断的文化差异，而中西方人所用的社交媒体并不相同，因此单独使用同一种平台是不够真实的。另外，无论是在微博还是 Twitter 上对于本文采用的不道德行为情景(如偷税漏税、侵犯版权等)的披露是不受限制的。

意见 12：实验中使用的道德责任的测量问题“Rico 应该承担多少指责”这个翻译不符合一般中文的语言习惯，翻译为“多大程度上归咎于 Rico”可能更合适。

回应：感谢您的建议。我们在新补充的实验 4 中根据您的建议将表述更换为“多大程度上应该归咎于 Rico”。此外，我们还重新补充了实验 2，采用新的道德责任量表来测量(Cameron et al., 2010; Cronbach’s $\alpha = 0.81$)，复制出了实验 1 的结果，说明我们得到的效应并不是因为测量条目的表述问题引起的。

意见 13：实验 2 中图 4（实验 3 的结果相似）的结果也可以理解为，中国被试认为真人和虚拟人没有区别(4.57, 4.67)，而英国被试则认为虚拟人的心智远远低于真人(3.62, 5.16)。也就是说这个交互作用实际上是由西方被试的差异来推动的。①是的确中国被试认为真人和虚拟人在心智和在道德责任上都无差别吗？②是否对中国的被试进行了注意力检查和操纵检验，虚拟组的被试是否真的注意到里面是虚拟人？③如果是的话，假设 1 的提出可能修改为以下假设更合理：“与西方消费者相比，中国消费者更倾向于认为虚拟人需要承担类人的道德责任”。

回应：感谢您的问题。①我们同意您的意见，我们所得到的研究结果也可以理解为在中国文化下，人们对虚拟人和真人的心智能力评价和道德责任判断更加相似(但并非没有差别)。我们理论的出发点是文化差异，且更多关注交互效应的结果，而不是比较真人和虚拟人的简单效应；②我们同意操纵检验可能是一个问题。因此，我们增加了一个研究(实验 2)来考察这一点。我们加入了操纵检验，验证了操纵的有效性在两个文化中没有区别。这也说明我们的操纵材料是有效的。③我们确实研究重点在于文化差异的调节效应上，已按照您的建议修改了假设 1 的表述。

意见 14：另外，心智能力的量表考察了不同的维度/方面，如果分不同的维度去看，东西方比较的结果如何？到底是哪个方面对道德责任起决定作用，是否和分维度的结果吻合（见上面的问题 7）？挖掘出背后更深层的原因无论是在理论和实践上都很有价值和意义。

回应：感谢您的问题。已有研究主要将**心智能力划分为自主力和感知力两个维度**(Gray et al., 2007)。在修改后的实验里，我们分别从这两个维度进行了分析，并且发现了感知力比自主力具有更强的解释力度，这与近期的相关研究也是吻合的(Dietze & Knowles, 2021)。而感知力的评价与类人的外观、社会交互都密切相关(Appel, 2012; Gray et al., 2011; Krumhuber et al., 2015)，这也是虚拟人区别与人工智能的重要因素之一。

意见 15：实验 3 目的是验证心智能力的中介作用。虽然作者排除了虚拟人熟悉程度和事件严重程度的作用。但是，存在其他更为可能和重要的替代性的中介变量，例如思考方式、归因、对 intentionality 的推断等。需要更多的实证证据来排除这些可能性的替代解释。

回应：我们非常赞同你的看法，也认为思考方式、归因、和对意图的推断是可能的解释。事实上，心智能力本身就包含了意图、计划和发挥自我控制的能力(Gray et al., 2011)，在我们对于心智能力的测量中已经包括了这些方面的条目。另外，有研究认为，思考方式和归因方式，例如换位思考能力、分析型思维和整体性思维、内外部归因，也会影响人们对心智的感知(Dietze & Knowles, 2016; Dietze & Knowles, 2021; Hackel et al., 2014)。因此，我们认为心智感知是最核心的中介变量，而思考方式和归因的文化差异是通过影响心智感知从而影响到道德责任判断的。

意见 16：实验 4 的实验设计为什么只有虚拟人，而没有包括真人的情境？

回应：感谢您的问题。因为我们的核心研究问题在于探讨中西方文化差异对社交媒体虚拟人道德责任判断的影响，前面的实验加入真人组作为控制组来验证虚拟人的特殊性。通过前面四项（更新后有五项）实验验证在真人组上不存在文化差异，并且中国人对虚拟人和真人的道德责任判断也不存在差异之后，在第四项（更新后的第五项）后续的行为实验中就不再考虑真人组的效应了。

意见 17：实验 4 所用材料中对虚拟人的背景介绍令人不解。虚拟人到底是真人的虚拟形象，还是基于人工智能的由公司或组织主导的数字实体？如果是后者，为什么会有如下信息“出生于 2002 年 3 月 21 日”？①在实践中两者都是存在，但是人们对这两种虚拟人的理解和判断很可能存在巨大差异。②例如，作者在文献、逻辑推导和管理启示的讨论中多以人工智能的相关研究为主，而实验中的操纵又是以基于真人的虚拟人（非公司、智能技术控制），存在着不一致。被试对虚拟人的理解是否与实验者的假定一致？需要对以上问题做出解释。

回应：感谢您的建议让我们进行了必要的调整和修改。

①现实生活中的虚拟人最经常出现的场景是社交网站，其中大多不会提到这个虚拟人到底背后是真人还是人工智能，是由公司操纵的还是由人工智能自动化的。因此我们的实验材料也是模拟了这种现实生活的场景，也就没有对其背后的运作机制进行具体的介绍。虽然现在大部分的虚拟人背后都是真人，但是随着技术的进步，或许还会出现完全由人工智能来控制的虚拟人。我们没有特别去强调虚拟人的运行原理是什么，因为大多数的社交媒体上遇到的虚拟人也都是避免提到这一点的。为了更加靠近真实生活当中人们需要凭直觉做出的道德判断，我们就采用了现在的实验材料。具体说来，我们给虚拟组的被试展示了虚拟偶像的定义：虚拟偶像是利用计算机图像软件制作的数字形象，他们以第一人称视角看待世界，并在媒体平台上占据一席之地；其次，我们借鉴了生活中虚拟偶像的描述，例如洛天依、默默酱等，例如“洛天依(Luo Tianyi)，7 月 12 日出生”，为了更加贴近生活，我们也加入了生日等信息的描述。

②我们同意您的看法，人们对于这两种虚拟人的理解和判断存在差异。由于我们的核心主效应在于文化差异，在不同文化中，被试看到同样的实验材料。如果不同的个体在理解和判断上存在差异，那么这种差异应该在两个文化当中也是类似的。就这一点，我们也增加了一个新的研究(实验 4)来考察这种理解是否存在影响。结果发现，无论虚拟人背后的主体是真实人类还是人工智能，中国人都会比西方人给出更高的道德责任评价。因此，我们的主要结果

并不是由于这种不同的理解来引发的。并且我们在之前的实验中排除了对虚拟人熟悉度的影响。也就是说，对虚拟人的理解的差异不会影响到我们的主要效应。由于目前对虚拟人的研究比较少，而且人们倾向于把虚拟人跟人工智能联系在一起，另外，人工智能本身也是一种类人的虚拟实体(Anderson et al., 2010)，因此我们在引言部分更多的借鉴了人工智能的研究来进行理论推导，详细内容放在“1.1 虚拟人的定义及研究范畴”中进行解释。

意见 18：管理启示需要进一步加强，目前有的不切题，有的观点过于片面，应围绕研究结论并联系实际来展开。

回应：感谢您的建议。由于本文重点在文化差异及道德判断理论上的贡献，因此将管理启示部分纳入到“7.2 对虚拟人的道德责任判断”部分的最后一段来陈述，并按照您的建议进行了修改。

审稿人 3 意见：

意见 1：我最大的疑问在于作者对于结果的解释。从实验 1（图 3）和实验 2（图 4）的结果来看，我认为更准确的表述并非是“东方人认为虚拟人具有更高的心智能力并需要承担更多的道德责任”，而应该是“西方人认为虚拟人具有更低的心智能力，因此需要承担更少的道德责任”。或者换言之，中国人认为虚拟人和真人没有什么不同（均值无显著差异），而西方人认为虚拟人不是“人”（均值与其他三组相比明显更低）。①造成这一结果可能的原因或许是中国人在看待虚拟人的时候更多地联想到了他们背后的实际操纵者（平台/团队/集团），基于背景和整体认知因而感知到了更多的主体性；②亦或仅仅是由于实验材料的原因（虚拟人和真人都是白人形象）导致中国人在比较虚拟人和真人的时候，由于他们都是外群体而没能感受到他们之间的差异。对于后面这个原因可能导致的问题，我希望作者能够补充说明，即为何在材料上都选取了白人的形象。

回应：我们同意您的解释，这主要取决于我们把什么看成是正常。如果我们认为正常的做法是把虚拟人看成跟自然人一样，那么这个解释更加贴切，应该是“西方人认为虚拟人具有更低的心智能力”。如果我们认为正常的做法是不认为虚拟人跟自然人一样，那么可能更贴切的表达是“中国人认为虚拟人存在更高的心智能力”。采纳您的意见，我们现在统一修改为：“中国人(比西方人)认为虚拟人具备更高的心智能力”。这样就重点考察虚拟人上面的文化差异。前人研究发现东方人比西方人更加多地进行换位思考(Wu & Keysar, 2007)，并且更加倾向于对其他事物进行拟人化(Letheren et al., 2016)。正因如此，我们推理出中国人会更加倾向于对虚拟人拟人，所以会把真实人类所特有的心智能力归因到虚拟人上面，从而导致了道德判断的文化差异。正如您所说，我们的研究也反应了，中国人对虚拟人的道德判断与真人更相似，而西方人认为虚拟人不是人，不具备人的心智能力，因此不需要承担道德责任。造成这一结果的原因在于心智感知的文化差异，我们的中介检验也一再的证明了这一点。

①中国人是不是在看虚拟人的时候更多地联想到了背后的实际操纵者呢？从我们的数据结果来看并不支持这一点。如果中国人更多联想到背后的操纵主体，那么按理中国人不应该认为虚拟人有更高的自主力(心智能力的一个重要维度)。逻辑上推理，应该联想到背后的操纵主体越多，对虚拟人本身的心智能力判断应该更低。为了回应您的问题，我们补充了一项新的实验来操纵虚拟人背后的主体。但结果发现，无论虚拟人背后是真实人类还是人工智能，中国人(比起西方人)都认为其需要承担更大的道德责任，因此从操纵主体的角度并不能解释文化差异的效应。

②那么是实验材料的原因吗？我们的结果也不支持这个结论。已经有大量研究表明，人们对外群体会有更多的去人性化，也就是我们会觉得外群体更加不像人(Cortes et al,, 2005; Kozak et al., 2006; Leyens et al., 2000; Schroeder & Epley, 2020)。在我们的实验中，中国人看到真人和虚拟人更像是外群体的白人，西方人看到的则是内群体的白人。那么从拟人的角度来说，中国人应该觉得白人的虚拟人更加不像人，因为它从人种上来说是外群体。西方人应该觉得白人的虚拟人更像人因为人种上是内群体。但是我们的结论刚好是反过来的，东方人觉得虚拟人更像人(即使是白人的虚拟人)，西方人觉得虚拟人更不像人(即使是内群体的白人虚拟人)。我们之所以选择白人作为实验材料也正是为了避免这个问题，所以选择了对我们的假设更加不利的白人实验材料。如果选择亚洲人的实验材料，那么就无法区分出来到底是虚拟人的效应还是内群体外群体的效应。

意见 2：在样本选择上，作者需要说明选取样本大小的理由，四个实验是否是四批不同的被试？为何实验 2 和实验 3 选取了 199 名英国被试而不是 200 名被试？是否存在被试损耗？这些问题需要加以说明。

回应：感谢您的建议，我们已在“实验 1”中的“2.1 实验设计与样本”第一段中对样本量的确定标准给出说明：“实验采用 G*Power 3.1 软件(Faul et al., 2007)计算实验所需样本量。对于本实验适用的双因素方差分析，当显著性水平 α 为 0.05 且效应量为中等效应时(f = 0.2 到 0.25 之间)，要达到 95%统计检验力所需要的样本量在 279 到 434 之间。为保证实验具有足够的样本量且为统一标准，本研究的所有实验按照每个条件下 100 名被试的标准招募，且采用原始数据进行分析。但由于线上平台存在同时作答以及未完成作答等情况，会出现多于或少于既定被试数量 1-2 个的情况。”

意见 3：实验 1 的道德情境是关于网暴，实验 3 的材料关于偷税漏税，实验 4 关于抄袭，为何要选择这些道德情境？这些问题需要加以说明。

回应：感谢您的建议，我们在“1.5 道德责任判断的文化差异及感知心智能力的中介作用”的第三段中给出了说明。这些道德情境是根据真实性来选择的。我们选取这四种情景(网络暴力、偷税漏税、侵犯版权和抄袭作品)都是属于真人与社交媒体虚拟人都可能做到的不道德行为，在被试看来会相对更加真实可信，并且更加贴近人们的生活，针对这些现实中频发的不道德行为也能够提供一定的管理启示。因为虚拟人不可能会涉入所有的不道德情境，比如虚拟人不可能杀人，因此只能选择最能让人信服的情境进行研究。而且为了外部效度，我们还采取了多个不同的道德情境，这样让我们的研究不只是局限于某一个情境。

意见 4：实验 3 为何要控制虚拟人熟悉度和事件严重性这两个变量需要作出一定的解释和说明。

回应：非常感谢您的建设性意见。根据您的建议，我们在实验 3 的第一段中已补充说明。在实验 3 中，我们排除了对虚拟人熟悉度和事件严重程度这两个替代性解释，原因是我们研究的核心在于中西方文化差异。我们发现相比于西方人，中国人认为虚拟人需要承担更大的道德责任，原因在于中国人对虚拟人心智能力的评价更高。但还有可能的解释是中国人比西方人更熟悉虚拟人，从而可能会对身边熟悉的东西更加倾向于拟人化。还有一种可能的解释是，中国人比西方人认为不道德情景中的事件(如偷税漏税)更严重，而导致在道德责任判断上的文化差异。因此，为了检验这两种替代性解释，我们测量了对虚拟人熟悉度和事件严重程度，通过中介模型的数据分析排除了这两种解释机制。

……

审稿人 4 意见：

意见 1：虚拟人不道德行为。本研究中选择了网络暴力、抄袭问题、偷税漏税不道德行为。那么，相对于网络暴力、抄袭问题，是否阅听人能感受到虚拟人会从事偷税漏税不道德行为？是否可能存在不同影响。

回应：感谢您的建议，我们非常同意您的看法。对于不同的道德事件，人们相信的程度会不一样。我们把这个可能性放到“7.3 研究局限与未来研究方向”第四段末进行讨论。在我们的研究中，这四种不道德行为都出现了同样的效应，因此即使被试不相信虚拟人可能偷税漏税，也不会影响到我们在文化差异上的结果。未来的研究也可以考察其他的一些不道德行为。

意见 2：对于东方人而言（中国），在心智能力感知，道德责任等对真人和虚拟人没有差异，主要差异来自于西方人，会否西方人更认为虚拟人不是人，本来就是假的，所以没必要对其进行道德判断，道德判断实际上主要与真实的人类活动相关。同上第 1 点，特别是如果其道德行为本身也和虚拟人行为连接较弱？而相对的，网络暴力、抄袭问题似乎与虚拟人行为连接或许相对较高。

回应：感谢您的建议。我们同意您的看法，正是因为西方人对虚拟人的拟人化程度更低，所以他们不认为虚拟人具备心智能力，从而没有必要对其进行道德责任的归因。您在这里提到的“假”其实已经反映在心智能力的评价之中，心智能力本身包含了意图、计划和发挥自我控制的能力(Gray et al., 2011)。如果西方人认为虚拟人假，则不会具备这些能力，从而在道德责任的评价上也相对更低，因此我们提出的机制同样可以解释这一观点。

意见 3：理论和假设部分：虚拟实体的心智感知与道德责任之间的方向及关系论述可加强。

回应：您的建设性意见让我们对理论和假设部分进行了更为深入的思考。我们已重新修改完善了这部分内容，详见“1.3 心智感知理论”及“1.5 道德责任判断的文化差异及感知心智能力的中介作用”部分。

意见 4：实验部分：仅在实验 4 实验中测量社经地位，是否存潜在影响？如存在潜在影响，可简单论述。

回应：感谢您提出的问题。在原文章的实验 1 到实验 3 中，我们没有测量社会经济地位。在原文章的实验 4(现文章实验 5)中，我们希望能够更加广泛的考虑可能会影响的人口统计变量，因此加入了社会经济地位和受教育程度的测量。根据您的建议，我们在原文章实验 4(现文章实验 5)的数据分析当中不论包不包括主观社会经济地位和受教育程度，结果都不发生改变。我们还新加了“附录三”，呈现了本文中所有实验加入控制变量后的结果，在控制性别和年龄等变量后都不改变结果的显著性和方向。另外如果您说的潜在影响是指的测量社经地位对于其他变量测量的影响的话，我们对社会经济地位等人口统计变量的测量都是放在问卷的最后一部分，因此也应该不会存在这个影响。

意见 5：注意内文引用及参考文献格式的统一及规范。

回应：感谢提醒，已全文检查并修正。

参考文献

Appel, J., Von Der Pütten, A., Krämer, N. C., & Gratch, J. (2012). Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Advances in Human−Computer Interaction*, *2012*, 1–10.

Anderson, E. F., McLoughlin, L., Liarokapis, F., Peters, C., Petridis, P., & De Freitas, S. (2010). Developing serious games for cultural heritage: A state-of-the-art review. *Virtual Reality*, *14*(4), 255–275.

Arsenyan, J., & Mirowska, A. (2021). Almost human? A comparative case study on the social media presence of virtual influencers. *International Journal of Human−Computer Studies*, *155*, 102694.

Chen, M. J., Wang, Y. S., Zhao, B. J., Li, X., & Bai, X. J. (2022). The role of text familiarity in Chinese word segmentation and Chinese vocabulary recognition. *Acta Psychologica Sinica*, *54*(10), 1151–1166.

[陈茗静, 王永胜, 赵冰洁, 李馨, 白学军. (2022). 中文文本熟悉性在词切分和词汇识别中的作用. *心理学报*, *54*(10), 1151–1166.]

Chen, Z. X., He, Y., Li, X., & Wang, L. (2022). Can you perceive my efforts? The impact of social status on consumers' preferences for complexity. *Acta Psychologica Sinica*, *54*(9), 1106–1121.

[陈增祥, 何云, 李枭, 王琳. (2022). 你能看见我的努力吗: 社会地位感知对消费者繁简偏好的影响. *心理学报*, *54*(9), 1106–1121.]

Cortes, B. P., Demoulin, S., Rodriguez, R. T., Rodriguez, A. P., & Leyens, J. P. (2005). Infrahumanization or familiarity? Attribution of uniquely human emotions to the self, the ingroup, and the outgroup. *Personality and Social Psychology Bulletin*, *31*(2), 243–253.

Dietze, P., & Knowles, E. D. (2016). Social class and the motivational relevance of other human beings: Evidence from visual attention. *Psychological Science*, *27*(11), 1517–1527.

Dietze, P., & Knowles, E. D. (2021). Social class predicts emotion perception and perspective-taking performance in adults. *Personality and Social Psychology Bulletin*, *47*(1), 42–56.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Feng, W. T., Xu, A. P., Huang, H., & Wang, T. (2022). Kawai vs. Whimsical: The influence of cuteness types of luxury brands on consumers' preferences. *Acta Psychologica Sinica*, *54*(3), 313–330.

[冯文婷, 徐瑗苹, 黄海, 汪涛. (2022). 萌萌哒还是古灵精怪? 奢侈品品牌可爱风格对消费者偏好的影响. *心理学报*, *54*(3), 313–330.]

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619−619.

Gray, K., Knobe, J., Sheskin, M., Bloom, P., & Barrett, L. F. (2011). More than a body: Mind perception and the nature of objectification. *Journal of Personality and Social Psychology*, *101*(6), 1207–1220.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101−124.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108.

Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. *Journal of Experimental Social Psychology*, *52*, 15-23.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*(4), 613–628.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, *10*(3), 252−264.

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, *65*(1),

399−423.

Hume, D. (1751). *An enquiry concerning the principles of morals*. Clarendon Press.

Kim, D., & Jo, D. (2022). Effects on co-presence of a virtual human: A comparison of display and interaction types. *Electronics*, *11*(3), 367.

Kozak, M. N., Marsh, A. A., & Wegner, D. M. (2006). What do I think you're doing? Action identification and mind attribution. *Journal of Personality and Social Psychology*, *90*(4), 543–555.

Krumhuber, E. G., Swiderska, A., Tsankova, E., Kamble, S. V., & Kappas, A. (2015). Real or artificial? Intergroup biases in mind perception in a cross-cultural perspective. *PLoS One*, *10*(9), e0137840.

Lee, H., Jung, T. H., tom Dieck, M. C., & Chung, N. (2020). Experiencing immersive virtual reality in museums. *Information & Management*, *57*(5), 103229.

Letheren, K., Kuhn, K.L., Lings, I., & Pope, N. Kl. (2016). Individual difference factors related to anthropomorphic tendency. *European Journal of Marketing*, *50*(5/6), 973–1002.

Leyens, J. P., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups. *Personality and Social Psychology Review*, *4*(2), 186–197.

Loveys, K., Sagar, M., & Broadbent, E. (2020). The effect of multimodal emotional expression on responses to a digital human during a self-disclosure conversation: A computational analysis of user language. *Journal of Medical Systems*, *44*(9), 1–7.

Moustakas, E., Lamba, N., Mahmoud, D., & Ranganathan, C. (2020, June). Blurring lines between fiction and reality: Perspectives of experts on marketing effectiveness of virtual influencers. In *2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)* (pp. 1–6). IEEE, Dublin, Ireland.

Russell, S., & Norvig, P. (2002). Artificial intelligence: A modern approach. New York: Prentice Hall.

Schroeder, J., & Epley, N. (2020). Demeaning: Dehumanizing others by minimizing the importance of their psychological needs. *Journal of Personality and Social Psychology*, *119*(4), 765−791.

Sun, B., Zeng, X. Q., Xu, K. Y., Xie, Y. T., & Fu, S. M. (2022). Kawai vs. Whimsical: The influence of cuteness types of luxury brands on consumers' preferences. *Acta Psychologica Sinica*, *54*(8), 867–880.

[孙博, 曾宪卿, 许恺煜, 谢韵婷, 傅世敏. (2022). 情绪面孔的意识神经相关物及其无意识自动加工: 来自事件相关电位的证据. *心理学报*, *54*(8), 867–880.]

Sung, E. C. (2021). The effects of augmented reality mobile app advertising: Viral marketing via shared social experience. *Journal of Business Research*, *122*, 75–87.

Sung, E. C., Han, D. I. D., Bae, S., & Kwon, O. (2022). What drives technology-enhanced storytelling immersion? The role of digital humans. *Computers in Human Behavior*, *132*, 107246.

Volante, M., Babu, S. V., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., ... Fasolino, T. (2016). Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE Transactions on Visualization and Computer Graphics*, *22*(4), 1326–1335.

Wang, L. L., & Dong, M. L. (2022). Does "male beauty" really work: The impact of male endorsements on female consumers' evaluation of female-gender-imaged product. *Acta Psychologica Sinica*, *54*(2), 192–204.

[王丽丽, 董梦璐. (2022). "美男诱惑"真的奏效吗: 男性代言女性产品对女性消费者产品评价的影响. *心理学报*, *54*(2), 192–204.]

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383−388.

Ward, A. F., Olsen, A. S., & Wegner, D. M. (2013). The harm-made mind: Observing victimization augments attribution of minds to vegetative patients, robots, and the dead. *Psychological Science*, *24*(8), 1437−1445.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, *18*(7), 600−606.

Xu, L. Y., Yu, F., & Peng, K. P. (2022). Algorithmic discrimination causes less desire for moral punishment than human discrimination. *Acta Psychologica Sinica*, *54*(9), 1076−1092.

[许丽颖, 喻丰, 彭凯平. (2022). 算法歧视比人类歧视引起更少道德惩罚欲, *心理学报*, *54*(9), 1076−1092.]

第二轮

审稿人 1 意见：

意见 1：中西方文化差异是一个非常大的命题。既然文章将“中西方文化差异”作为一个重要自变量，那至少要对“中西方文化差异”作个简要综述。中西方被试为什么对虚拟人的态度不一样，其深层的文化心理原因是什么？文章中没有出现文化心理学方面的理论，也没有用跨文化心理学已有的经典文献来支撑或解释这种现象。甚至经典的跨文化心理学研究作者和文献都没有或很少出现。因此，要突出中西方文化差异的影响，综述部分和讨论部分应该增加文化的理论厚度。

回应：感谢您的建议。我们在本文的综述部分增加了一小节“1.1 中西方文化差异对道德判断的影响”，补充加入了文化心理学方面的理论以及跨文化心理学已有的经典文献，增加了文化的理论厚度，具体内容如下所示。此外，我们重新修改了本文的综合讨论部分，增加了文化相关的元素和厚度，详细内容见“意见 4”的回应部分。

1.1 中西方文化差异对道德判断的影响

文化会对人们的心理与行为产生重要影响(侯玉波, 朱滢, 2002)。文化心理学是运用心理学的理论和方法来研究文化对人类心理功能的影响(刘邦惠, 彭凯平, 2012)。在文化心理学中，文化泛指社会成员之间所共有的价值观、规范、思维方式和行为方式等(Hofstede, 1980; Morling, 2016; Na et al., 2010; 黄梓航 等, 2018)。Hofstede (1980)提出，文化价值包括个人主义/集体主义、权力距离、不确定性规避和刚柔性四个维度。其中个人主义/集体主义成为文化心理学中应用最广泛、影响最大的一个文化维度，引发了大量的后续研究(Greenfield, 2009; Oyserman et al., 2002)。个人主义文化强调人们的独立性、独特性和自由选择，集体主义文化则强调人们的互依性、社会嵌入性和对内群体的义务和忠诚(Grossmann & Na, 2014; Oyserman et al., 2002; 黄梓航 等, 2018)。个人主义/集体主义的文化心理学理论认为，心理上的文化差异来源于不同文化的价值取向。西方文化的价值取向具有较强的个人主义意识，而东方文化的价值取向具有较强的集体主义意识(刘邦惠, 彭凯平, 2012)。例如，Morris 和 Peng (1994)的研究表明，美国人的归因方式是以个人为中心，而中国人的归因方式是以集体为中心。在跨文化心理学领域，许多关于东西方文化差异的研究都是基于个人主义/集体主义的文化理论(Triandis, 1989)。

先前的研究发现，东西方文化差异会影响人们的道德判断。西方文化强调以个人自由、权利、公正、关爱和宽恕等为导向的道德观，而东方文化强调以集体和个人责任为导向的道德观(彭凯平等, 2011; 王恩界, 乐国安, 2006)。Miller 和 Bersoff (1992)的研究指出，西方人更强调公正伦理，而东方人更强调责任伦理。个人主义/集体主义文化也会影响人们的道德决策，因为它们涉及到个人利益还是集体利益优先的信念(Oyserman et al., 2002)。在本文中，基于个人主义/集体主义的文化心理学理论，我们关注到了中西方文化差异对虚拟人道德责任判断的影响。与个人主义文化(如美国)相比，集体主义文化(如中国)下的人们会有更高的换位思考能力(Wu & Keysar, 2007)和拟人化倾向(Letheren et al., 2016)。那么，相比于西方文化，中国文化影响下的人们是否会对虚拟人的心智能力评价更高，进而认为虚拟人应该为不

道德行为承担更大的道德责任？这是本文想要回答的问题。

意见 2：文章将心智感应作为文化差异的一个重要中介变量，那么应考虑怎样从文化差异引出心智感应。其次从作者的综述来看，心智感应差异仍然只是一种现象，而非一种文化本质，还是没有深入到文化差异的理论或机制。关于这块的引入和综述，希望作者能挖掘其深层的文化理论。正如审稿人一的意见“如果要考察的是对虚拟人-自然人之间的道德责任判断差异，那么引入心智感知还算合理；但如果要考察的是文化差异，为什么要考虑心智感知作为文化差异的心理机制，是不是过于牵强了？可以说揭示的仍是现象，没有深入到机制。”从作者回应及文章修改来看，作者在这方面还需要深入。

回应：感谢您的建议。在本文的综述部分，我们增加了一小节“1.1 中西方文化差异对道德判断的影响”，补充加入了文化心理学方面的理论以及经典文献。具体来说，在 1.1 的第二段最后，我们提到“在本文中，基于个人主义/集体主义的文化心理学理论，我们关注到了中西方文化差异对虚拟人道德责任判断的影响。与个人主义文化(如美国)相比，集体主义文化(如中国)下的人们会有更高的换位思考能力(Wu & Keysar, 2007)和拟人化倾向(Letheren et al., 2016)。那么，相比于西方文化，中国文化影响下的人们是否会对虚拟人的心智能力评价更高，进而认为虚拟人应该为不道德行为承担更大的道德责任？这是本文想要回答的问题。”此外，在修改后的“1.5 心智感知及道德责任判断的文化差异”中，我们解释了如何从文化差异引出心智感应差异，增加了文化的理论厚度。具体来说，在 1.5 的第一段，我们提到“已有文献通常会从感知者的角度探讨心智感知的前因(Waytz et al., 2010)，并且认为文化差异是影响心智感知的重要因素(Dietze & Knowles, 2021)。一方面，文化差异会影响人们的换位思考能力。例如，比起个人主义文化(如美国)，集体主义文化(如中国)背景下的人们会表现出更高的换位思考能力(Wu & Keysar, 2007)。比起英国人，中国人对人工智能机器人社会合作属性的感知更高(Dang & Liu, 2022)。还有研究发现，低社会阶层(相比于高社会阶层)的人们往往表现出更多的社会互动和更高的换位思考能力(Dietze & Knowles, 2016)，心智感知能力也更高(Dietze & Knowles, 2021)。因此，比起个人主义文化影响下的西方人，集体主义文化下的中国人更可能将自己的情感和能力投射到别人的身上，可能会对虚拟人的心智能力评价更高。另一方面，文化差异会影响人们的拟人化倾向。例如，Letheren 等(2016)发现，东亚人比白种人具有更高的拟人化倾向。也有研究指出，西方文化的基督教和伊斯兰教认为神更不具有固定形状的身体，而东方文化的佛教和印度教认为神具有有形的身体(Fuller, 2004; Ohnuma, 2007; Samuel, 1989; Willard & McNamara, 2019)。东方人不但会给神赋予一个人的身体，还会给神赋予人类特有的能力(McGuire, 2018)。现有的研究指出，拟人化倾向会影响心智能力的感知，拟人化倾向越高的被试对机器模型 Pleo 心智能力的感知越高(Eyssel & Pfundmair, 2015)。因此，东方人比西方人更加倾向于对非人类实体进行拟人化，可能会更加认为虚拟人具备类似真人的心智能力。综上，基于个人主义/集体主义的文化心理学理论，本文认为，比起西方文化，中国文化影响下的人们对虚拟人的心智能力评价会更高。”

意见 3：各个实验后的小讨论，建议适当增加文献的对比，而不是简单的结果陈述。

回应：感谢您的建议。在各个实验后的小讨论中，我们适当地增加了文献的对比，对小讨论部分进行了完善，具体内容如下所示。

2.4 讨论

实验 1 采用曝光隐私引发网络暴力的不道德行为情景来检验中西方文化对虚拟人作为道德主体在道德责任判断上的差异。结果显示，当人们看到虚拟人的不道德行为后，相比于西方文化，中国文化影响下的人们认为其需要承担更大的道德责任。这一结果从文化差异的角度拓展了人们对虚拟实体道德责任归因的文献(Awad et al., 2020; Malle et al., 2015; Young

& Monroe, 2019; 褚华东 等, 2019)，表明不同文化背景下的人们对虚拟人道德责任的判断存在差异。

3.4 讨论

实验 2 通过新的不道德行为情景(偷税漏税)和道德责任测量条目(Cameron et al., 2010)表明，比起西方文化，中国文化影响下的人们认为虚拟人在出现偷税漏税的行为后需要承担更大的道德责任。实验 2 复制了实验 1 的结果，扩展了与虚拟实体道德责任相关的研究(Awad et al., 2020; Malle et al., 2015; Young & Monroe, 2019; 褚华东 等, 2019)。在实验 3 中，我们将考察对虚拟人感知心智能力的中介机制。

4.4 讨论

综合以上结果，实验 3 首先通过有调节的中介检验，验证了感知心智能力的中介作用。结果表明，比起西方文化，中国文化影响下的人们认为虚拟人具有更高的心智能力，因此需要承担更大的道德责任；而在真人上没有显著的中西方文化差异。这一发现进一步表明文化差异对心智感知存在重要影响(Dietze & Knowles, 2021)。其次，实验 3 通过将心智能力中的认知自主力和情绪感知力同时纳入模型进行平行中介分析，发现这一过程由情绪感知力而非认知自主力完全中介，验证了情绪感知力存在更强的解释力度(Gray & Wegner, 2012)。最后，实验 3 还排除了对虚拟人的熟悉度、事件严重性的替代性解释。在接下来的研究中，我们将考察是否不同类型的虚拟人会影响人们对他们的道德判断。

5.4 讨论

实验 4 通过操纵虚拟人背后的主体类型，分别在真人组、虚拟真人组和虚拟人工智能之间比较中西方文化差异。结果发现，无论虚拟人背后的主体是真实人类还是人工智能，相比于西方文化，中国文化影响下的人们都会赋予其更大的道德责任，而在真人组不存在这种文化差异。这说明，道德责任判断的中西方文化差异同样存在于由真实人类和人工智能驱动的虚拟人当中。以往有研究认为，在对虚拟实体做出责任判断时会考虑其背后真实主体(Bryson et al., 2017; Champagne & Tonkens, 2015; Constantinescu et al., 2022)，但本实验结果发现虚拟人背后的主体类型并不会影响对其道德责任判断的文化差异。在实验 5 中，我们将考察人们得知虚拟人不道德行为后对人们后续行为的影响。

6.4 讨论

实验 5 首先通过抄袭作品的不道德情景验证了中西方文化差异对虚拟人道德责任判断的影响。相比于西方文化，中国文化影响下的人们认为虚拟人需要承担更大的道德责任，复制了前面实验的结果。其次，本实验检验了对虚拟人道德责任判断的文化差异所产生的的行为后果。相比于西方文化，中国文化影响下的人们认为虚拟人更应该被罚款并被封禁社交账号，同时更大比例的人们选择取消对虚拟人社交媒体账号的关注。最后，本实验通过连续中介模型验证了感知心智能力和道德责任判断的连续中介作用，进一步验证了本研究机制的稳健性。先前对虚拟实体道德责任的研究大多停留在态度评价和感知上(Awad et al., 2020; Malle et al., 2015; Young & Monroe, 2019; 褚华东 等, 2019)，实验 5 采用了贴近现实的不道德情景，并考察了针对虚拟实体的道德惩罚问题，在一定程度上丰富并延伸了虚拟实体不道德行为带来的后果。

意见 4：综合讨论部分，中西方文化差异是本文最主要探讨的问题，却没有用一节来进行讨论解释。“7.1 对虚拟人的心智感知”这只是手段而不是目的，文章的目的应该还是要回到文化差异上来。虽然讨论中多次提到文化二字，但到底是什么文化，是什么文化心理呢？通过系列实验研究，是支持了还是反对了哪个文化理论呢，这才是要讨论的重点。整体讨论仍然是再次描述实验结果或现象，而不是理论分析。

回应：感谢您的建议让我们对综合讨论部分进行了更好的完善。我们将原文的“7.1 对虚拟

人的心智感知”修改为“7.1 中西方文化差异对虚拟人心智感知的影响”；将原文的“7.2 对虚拟人的道德责任判断”修改为“7.2 中西方文化差异对虚拟人道德责任判断的影响”；将原文 7.2 的最后一段修改为“7.3 实践意义”。在修改后的综述讨论部分，我们从中西方文化差异的角度出发，补充加入了文化心理学方面的理论和文献，增加了讨论部分文化的理论厚度。具体内容如下所示：

7.1 中西方文化差异对虚拟人心智感知的影响

文化会对人们的心理与行为产生重要影响(侯玉波, 朱滢, 2002)。随着虚拟现实技术的发展以及元宇宙概念的兴起，越来越多的虚拟人出现在人们的日常生活中，然而目前国内外大部分有关虚拟实体的文献都是以人工智能、算法为研究对象，很少有研究关注到虚拟人。在本文中，我们关注到了中西方文化差异对虚拟人感知和行为反应的影响，在一定程度上能够弥补虚拟人相关文献的缺口。

本文拓展了文化差异和心智感知相关的文献，验证了中西方文化差异对虚拟人心智感知的影响。过去有关心智感知前因的研究大多探讨了感知者和被感知者的个体差异，例如感知者的社会需要动机(Waytz et al., 2010; Dietze & Knowles, 2016)、主观社会地位(Dietze & Knowles, 2021)、换位思考能力(Dietze & Knowles, 2016; Dietze & Knowles, 2021)和拟人化倾向(Eyssel & Pfundmair, 2015)等，以及被感知者的视觉外观(Appel et al., 2012; Gray et al., 2011; Krumhuber et al., 2015)、情绪表达(Gray & Wegner, 2012)和拟人化程度(Waytz et al., 2010)等。也有研究表明，文化差异是影响心智感知的重要因素(Dietze & Knowles, 2021)。例如，Willard 和 McNamara(2019)通过跨文化样本的问卷调查发现，北美人和斐济人对上帝心智能力的看法并不一致，斐济人认为上帝的心智能力和人类相似，而北美人认为上帝在感受维度上的心智能力更低于人类。最近一项研究表明，比起英国人，中国人对人工智能机器人社会合作的评价更高(Dang & Liu, 2022)，但该研究并没有测量不同文化背景下的被试对人工智能机器人的心智感知。为了弥补这一缺口，本文将虚拟人与真人做对比，研究了中西方文化差异对虚拟人心智感知的影响。结果表明，与西方文化相比，中国文化影响下的人们会赋予虚拟人更高的心智能力。

本文的研究发现在一定程度上支持了个人主义/集体主义的文化心理学理论。该理论认为，心理上的文化差异来源于不同文化的价值取向。西方文化的价值取向具有较强的个人主义意识，而东方文化的价值取向具有较强的集体主义意识(刘邦惠, 彭凯平, 2012)。在文化心理学的领域中，许多关于东西方文化差异的研究都是基于个人主义/集体主义的文化理论(Triandis, 1989)。例如，与个人主义文化(如美国)相比，集体主义文化(如中国)下的人们会表现出更高的换位思考能力(Wu & Keysar, 2007)和拟人化倾向(Letheren et al., 2016)。本文通过 5 个实验证明，与西方文化(美国/英国)相比，中国文化下的人们会认为虚拟人具备更高的心智能力。这一研究发现丰富了与个人主义/集体主义相关的文献(Letheren et al., 2016; Wu & Keysar, 2007)。

此外，本研究发现了人们对人类和虚拟人心智感知的差异。实验 3 的结果表明，无论是中国人($F(1, 395) = 6.85$, $p = 0.009$, $\eta_p^2 = 0.017$)还是美国人($F(1, 395) = 81.35$, $p < 0.001$, $\eta_p^2 = 0.171$)，都会认为虚拟人的心智能力比真人更低，这与人工智能、机器人的相关研究结果是吻合的(Gray et al., 2007; Broadbent, 2017)。因此，本文也将虚拟人和真人进行了对比，扩大了心智感知在人机交互中的研究范畴。

7.2 中西方文化差异对虚拟人道德责任判断的影响

本文对虚拟实体道德判断的研究进行了延伸，验证了中西方文化差异在对虚拟人道德责任判断中的影响。以往研究在对虚拟实体能否承担道德责任的问题上存在争议。根据经典道德责任理论，虚拟实体不具有意图和情绪，不能承担道德责任(Hakli & Mäkelä, 2019; Parthemore & Whitby, 2014)。然而，近年来的实证研究发现，当人工智能犯错时，人们依然

会将责任归咎于它们，只是与真人相比程度不同(Awad et al., 2020; Young & Monroe, 2019)。Coeckelbergh(2020)同时强调了责任主体和责任客体在对虚拟实体道德判断中的重要性，而心智感知中的认知自主力与责任主体、情绪感知力与责任客体的判断存在密切联系(Gray et al., 2007)。因此，本人引入文化差异的概念，将研究对象从人工智能拓展到虚拟人，研究中西方文化差异对虚拟人道德责任判断的影响，并验证了感知心智能力的中介机制，丰富了心智感知、道德判断与中西方文化差异的相关文献。

在探讨中介机制的研究中，本文将心智能力划分为认知自主力和情绪感知力两个维度进行了分析，发现了情绪感知力比认知自主力具有更强的解释力度，这与 Sullivan 和 Wamba (2022)的研究结论相吻合。他们也发现，当人工智能故意(相比于意外)伤害人类时，人们会更多责备人工智能本身，而这一效应由情绪感知力而非认知自主力中介。由于认知自主力包含了道德责任的归属的两个重要前因，即自由意志和认知(Aristotle, 1999)，一般认为认知自主力评价越高，越应该为行为后果承担道德责任(Gray & Wegner, 2009; Gray & Wegner, 2012)，但越来越多的研究发现情绪感知力评价在道德判断中的重要性(Bigman & Gray, 2018; Greene et al., 2001; Sullivan & Wamba, 2022)。因此，本文研究结论进一步支持了情绪感知力在道德责任判断中不可忽视的影响。与 Sullivan 和 Wamba (2022)的研究不同，本文的研究重点关注中西方文化差异对虚拟人道德责任判断的影响。本文的研究表明，相比于西方文化，中国文化影响下的人们认为虚拟人犯错后需要承担更大的道德责任。本文除了发现感知心智能力的中介外，还探索了虚拟人主体类型的调节作用以及中西方文化差异的后续影响（对虚拟人犯错后的道德惩罚）。

7.3 实践意义

本研究对于虚拟人的设计、运营和在道德伦理上的治理问题提供了一定的实践指导意义。由于近年来明星丑闻频频曝光，越来越多的企业选择与虚拟人合作进行营销活动，甚至创造自己企业的虚拟人，如屈臣氏的 AI 品牌代言人屈晨曦，普拉达香水代言人 Candy 等。那么，虚拟人真的不会翻车吗？由于技术和相关伦理问题，目前社交平台上的虚拟人依然由人类在背后操纵，无法实现真正意义上的自主，因此出现了虚拟人抄袭、侵权、强奸等不道德事件。然而，本研究发现不同文化背景下的消费者对虚拟人不道德行为的态度和反应会有差异。例如，在中国文化中，社交媒体上的虚拟人犯错后，不论背后主体是真实人类还是人工智能，人们都会认为虚拟人和真人会承担类似的道德责任，同时也会对虚拟人进行惩罚(如取消关注)。有研究者也认为赋予虚拟人等虚拟实体道德责任能够保证法律的连续性，确立虚拟实体在社会中的法人地位(Van Genderen, 2018)，因此相关部门需要针对此类问题制定更加完善的治理方针。另外，我们的研究结果能够帮助虚拟人的设计者和运营者了解虚拟人如何被人们评价和感知。由于本研究发现，情绪感知力的评价决定了道德责任的判断，因此当设计者赋予虚拟人更加拟人化的外观、情绪和社会互动时，运营者应更仔细地考量虚拟人的公共言论，并做出道德上可接受的行为表达，同时也要准备好面对相关的法律和伦理问题。

……

审稿人 2 意见：

意见 1：作者将心智能力(Mental Capacity)分为认知的自主力(Agency)和情绪的感知力(Experience)，后文则简化为自主力和感知力。在 1.2 的第二段也用自主力来指代自由意志，两个概念有所混淆。由于自主和感知两个概念的内涵比较宽泛，容易产生歧义或误解，建议作者在行文中不要简化，而是采用完整的“认知自主”和“情绪感知”来书写。

回应：感谢您的修改建议。根据您的指导，我们已将全文及附录中涉及到“自主力”和“感知力”的表述均修改为“认知自主力”和“情绪感知力”。

意见 2：1.4 和 1.5 的内容有诸多重复之处，显得冗余，建议将两个小节合并做必要的精简和提炼。

回应：感谢您的建议。我们已将“1.4 心智感知的文化差异”和“1.5 道德责任判断的文化差异及感知心智能力的中介作用”这两小节合并为“1.5 心智感知和道德责任判断的文化差异”，并做了必要的精简和提炼。具体内容如下所示：

1.5 心智感知及道德责任判断的文化差异

已有文献通常会从感知者的角度探讨心智感知的前因(Waytz et al., 2010)，并且认为文化差异是影响心智感知的重要因素(Dietze & Knowles, 2021)。一方面，文化差异会影响人们的换位思考能力。例如，比起个人主义文化(如美国)，集体主义文化(如中国)背景下的人们会表现出更高的换位思考能力(Wu & Keysar, 2007)。比起英国人，中国人对人工智能机器人社会合作属性的感知更高(Dang & Liu, 2022)。还有研究发现，低社会阶层(相比于高社会阶层)的人们往往表现出更多的社会互动和更高的换位思考能力(Dietze & Knowles, 2016)，心智感知能力也更高(Dietze & Knowles, 2021)。因此，比起个人主义文化影响下的西方人，集体主义文化下的中国人更可能将自己的情感和能力投射到别人的身上，可能会对虚拟人的心智能力评价更高。另一方面，文化差异会影响人们的拟人化倾向。例如，Letheren 等(2016)发现，东亚人比白种人具有更高的拟人化倾向。也有研究指出，西方文化的基督教和伊斯兰教认为神更不具有固定形状的身体，而东方文化的佛教和印度教认为神具有有形的身体(Fuller, 2004; Ohnuma, 2007; Samuel, 1989; Willard & McNamara, 2019)。东方人不但会给神赋予一个人的身体，还会给神赋予人类特有的能力(McGuire, 2018)。现有的研究指出，拟人化倾向会影响心智能力的感知，拟人化倾向越高的被试对机器模型 Pleo 心智能力的感知越高(Eyssel & Pfundmair, 2015)。因此，东方人比西方人更加倾向于对非人类实体进行拟人化，可能会更加认为虚拟人具备类似真人的心智能力。综上，基于个人主义/集体主义的文化心理学理论，本文认为，比起西方文化，中国文化影响下的人们对虚拟人的心智能力评价会更高。

心智感知理论将心智能力划分为认知自主力和情绪感知力两个维度(Gray et al., 2007; Gray et al., 2011; Gray & Wegner, 2009)。由于认知自主力涉及计划、思考、行动和自我控制等能力，对一个实体赋予更高的认知自主力意味着该实体能够作为道德主体做出道德决策并为行为负责(Gray et al., 2012; Gray & Wegner, 2009; Himma, 2009)。而研究也发现，涉及情绪体验的感知力对道德主体的判断同样至关重要(Greene et al., 2001; Sullivan & Wamba, 2022)。例如，自闭症和精神病患者缺少道德决策能力与情感体验息息相关(Bigman & Gray, 2018)。Himma (2009)认为，赞扬和谴责缺乏情绪感知力的实体是没有意义的。因此，对一个实体心智能力在两个维度上的评价都有可能影响人们对其道德责任的判断。

然而，目前对虚拟实体道德责任判断的文献存在分歧。部分学者认为人工智能等虚拟实体不能够承担道德责任(Hakli & Mäkelä, 2019; Parthemore & Whitby, 2014)。但也有研究表明，人们会将道德责任归因于虚拟实体(Awad et al., 2020; Malle et al., 2015; Young & Monroe, 2019; 褚华东 等, 2019)。为了得出一致的结论，需要深入探索人们对这些虚拟实体做出道德判断背后的原因。结合心智感知在道德责任判断中的重要作用，我们可以推断出，中西方文化差异能够通过影响心智能力的评价，来影响人们对虚拟实体的道德责任判断。

但目前针对虚拟实体道德责任的研究大多针对人工智能，并且采用道德两难问题的研究范式来进行实证检验(Awad et al., 2020; Malle et al., 2015; Young & Monroe, 2019; 褚华东 等, 2019)，而传统道德两难问题(如电车实验)的道德情景将理性与直觉剥离开，被认为是非典型的(Schein & Gray, 2018)。随着越来越多的虚拟人出现在人们的日常生活中，以虚拟实体为主体的道德问题不再局限于道德困境，取而代之的是更多类人的不道德行为，如侵权、强奸等。因此，本文聚焦社交媒体上的虚拟人贴近现实的道德情景，进行了实证研究，并提出以下研究假设：

假设 1：当看到虚拟人进行不道德行为后，中国文化(比西方文化)影响下的人们认为虚拟人需要承担更大的道德责任。
假设 2：中国文化(比西方文化)影响下的人们认为虚拟人具备更高的心智能力，这种更高的心智能力归因导致他们更加认为虚拟人应该为不道德行为承担道德责任。

意见 3：有些句子的意思不清或有误，需要重写。例如，1.4 在讲到文化差异时，“低社会阶层的文化会促进互依的价值观，通过社会关系和合作以适应资源匮乏的环境；而高社会阶层的文化会形成独立的行为策略(Kraus et al., 2012; Markus, 2017; Piff et al., 2012)”。①什么是低社会阶层的文化，这是国家间的文化差异吗？在 1.3 中作者提到“将自主力归因于一个实体意味着观察者认为该实体能够作为责任主体，像成年人一样行动、计划、实施自我控制、记忆、交流和思考；将感知力归因于一个实体表明观察者相信该实体能够作为责任客体，存在情绪感知能力，例如感到快乐、痛苦、愤怒等(Gray et al., 2007)。”该研究讨论的都是责任主体的问题，而感知力起到了重要的作用。②那么，感知力作为与责任主体或客体的关系到底是怎样的，需要进一步澄清。
回应：感谢您的问题和建议，我们重新修改并澄清了相关表述。

①我们同意您的看法，关于“低社会阶层的文化”和“高社会阶层的文化”的表达不太合适，应该改为“低社会阶层的人们”和“高社会阶层的人们”。在意见 2 的基础上，我们在对“1.4 心智感知的文化差异”的内容进行精简和提炼时，删去了这句话。

②已有的研究将心智能力划分为认知自主力和情绪感知力。分开来看时，认知自主力使得实体能够成为道德主体，情绪感知力使得实体能够成为道德客体(Gray et al., 2012; Ward et al., 2013)，都属于必要条件。例如，人们认为具有情绪感知力的实体在受到伤害才后会感受到疼痛，但并没有否认情绪感知力在道德判断形成中的作用。那么，在心智能力的整体归因中，情绪感知力是区分人类和其他实体的本质特征，通常被认为比认知自主力具有更强大的解释力(Gray & Wegner, 2012)。同样，心理学研究表明感受他人的痛苦而表现出同理心也是道德判断的核心要素之一(Bigman & Gray, 2018)。因此，情绪感知力也会影响对道德主体责任承担的判断。本文聚焦在研究人们对虚拟人作为道德主体时的道德责任判断。

意见 4：实验 2 选择用 0、1 变量来进行操纵的检验。为什么采用了这样的方法？结果发现，有一定比例的被试将真人认定为虚拟人，这些被试是否因为注意力等问题做出这样的回答，是否需要从被试中去除？如果去除的话，对结果是否有影响。
回应：感谢您的提问，如果我们采用 7 点量表来进行对博主类型的操纵检验(如 1 = 虚拟博主， 7 = 真人博主)会显得有些奇怪，因此直接采用了二分量表来操纵检验的测量。如果去除操纵检验不符合要求的被试，依然不会影响现有结果的方向和显著性。具体如下所示：

- 实验 2：排除不符合操纵检验的 17 名被试后，剩余 382 名被试，包括 189 名英国被试(69.84%女性; $M_{年龄}$ = 29.96 岁, $SD_{年龄}$ = 6.03 岁)和 193 名中国被试(65.28%女性; $M_{年龄}$ = 26.60 岁, $SD_{年龄}$ = 5.49 岁)。以文化类型(英国人编码为 0, 中国人编码为 1)和博主类型(真人编码为 0, 虚拟人编码为 1)为自变量，道德责任判断为因变量进行方差分析。结果显示，文化类型的主效应显著($F(1, 378) = 7.49$, $p = 0.006$, $\eta_p^2 = 0.019$)，博主类型的主效应显著($F(1, 378) = 27.02$, $p < 0.001$, $\eta_p^2 = 0.067$)，文化类型和博主类型的交互作用对道德责任判断具有显著影响($F(1, 378) = 9.67$, $p = 0.002$, $\eta_p^2 = 0.025$)。具体而言，比起英国人，中国人认为虚拟人应该承担更大的道德责任($M_{中国}$ = 5.37, SD = 1.19, 95%CI = [5.12, 5.62]; $M_{英国}$ = 4.60, SD = 1.67, 95%CI = [4.35, 4.86]; $F(1, 378) = 17.64$, $p < 0.001$, $\eta_p^2 = 0.045$)；但对真人的道德责任判断没有显著差异($M_{中国}$ = 5.64, SD = 1.07, 95%CI = [5.38, 5.90]; $M_{英国}$ = 5.69, SD = 1.04, 95%CI = [5.43, 5.95]; $F(1, 378) = 0.07$, $p = 0.796$)。此外，

中国人对真人和虚拟人的道德责任判断不存在显著差异($F(1, 378) = 2.21$, $p = 0.138$)。以上结果表明，排除不符合操纵检验的被试后，并不影响原有的结果。

- 实验 4：排除不符合操纵检验的 51 名被试后，剩余 549 名被试，包括 277 名英国被试(66.43%女性; $M_{年龄}$ = 30.98 岁, $SD_{年龄}$ = 5.34 岁)和 272 名中国被试(68.75%女性; $M_{年龄}$ = 26.86 岁, $SD_{年龄}$ = 5.82 岁)。以文化类型(美国人编码为 0, 中国人编码为 1)和博主类型(真人编码为-1, 虚拟真人编码为 0, 虚拟人工智能编码为 1)分别作为自变量，道德责任判断作为因变量进行方差分析。结果显示，文化类型的主效应显著($F(1, 543) = 82.08$, $p < 0.001$, $\eta_p^2 = 0.131$)，博主类型的主效应也显著($F(2, 543) = 23.38$, $p < 0.001$, $\eta_p^2 = 0.079$)，文化类型和博主类型的交互作用对道德责任判断具有显著的影响($F(2, 543) = 10.07$, $p < 0.001$, $\eta_p^2 = 0.036$)。具体而言，比起英国人，中国人认为无论虚拟人背后的主体是真实人类($M_{中国}$ = 5.27, SD = 1.33, 95%CI = [4.97, 5.57]; $M_{英国}$ = 3.75, SD = 1.47, 95%CI = [3.46, 4.05]; $F(1, 543) = 50.22$, $p < 0.001$, $\eta_p^2 = 0.085$)还是人工智能($M_{中国}$ = 5.11, SD = 1.20, 95%CI = [4.83, 5.39]; $M_{英国}$ = 3.77, SD = 1.57, 95%CI = [3.50, 4.05]; $F(1, 543) = 44.09$, $p < 0.001$, $\eta_p^2 = 0.075$)，都需要承担更大的道德责任；但对真人道德责任的判断没有显著的文化差异($M_{中国}$ = 5.48, SD = 1.32, 95%CI = [5.21, 5.75]; $M_{美国}$ = 5.14, SD = 1.31, 95%CI = [4.87, 5.41]; $F(1, 543) = 3.03$, $p = 0.082$)。此外，中国人对真人和两种类型的虚拟人的道德责任判断也没有显著差异($F(2, 543) = 1.70$, $p = 0.184$)，且两两比较的结果都不存在显著差异($ps > 0.187$)。

因此，在文章中两处操纵检验部分的最后，我们加了一句话：“此外，排除不符合操纵检验要求的被试并不会影响现有结果的方向和显著性。”

意见 5：全文 5 个实验虽然采用了不同的不道德行为，但是情境的设定都是一样的，都选择了女性的博主。这样单一的情境设定会也会在一定程度上影响结果的可推广性。而商业实践中的虚拟人类型是非常丰富多样的，如游戏人物、品牌代言人，虚拟偶像等等。建议采用其他的实验情境设定来 replicate 其中的部分实验。

回应：感谢您的建议。我们确实考虑到虚拟人类型的多样性，因此在实验 5 中采用的是虚拟偶像而非虚拟博主的描述和介绍。另外，由于目前国内外知名虚拟人通常是女性，因此本文采用的是女性虚拟人的实验材料。我们将这一局限性补充在综合讨论中，未来研究可以继续探索中西方文化差异对其他类型的虚拟实体的道德责任判断。具体来说，我们在“7.4 研究局限与未来研究方向”的最后第二段提到：“除此之外，未来研究还可以考虑其他调节变量，例如虚拟人的个性、外观及类型。”“另外，由于目前国内外知名虚拟人通常是女性，因此本文均采用了女性虚拟人的实验材料。但是，目前网络上涌现出许多不同类型的虚拟网红，例如动物型(如企鹅 Puff)、食物型(如蛋糕 Cupcake)等。未来研究可以进一步探讨中西方文化差异对男性虚拟人以及其他类型虚拟实体的道德责任判断。”

意见 6：文中提到了一篇近期的文献，Sullivan & Wamba (2022)的研究也提出并发现了心智能力的中介作用。需要说明和 Sullivan & Wamba (2022)研究相比，该研究的区别和贡献在哪里。

回应：感谢您的建设性意见。Sullivan 和 Wamba (2022)的研究发现，当人们认为人工智能故意(相比于意外)伤害人类(相比于非人类)时，人们会将更高的道德责任归咎于人工智能，这一效应是由心智能力中的情绪感知力而非认知自主力来中介的。与 Sullivan 和 Wamba (2022)的研究不同，本文重点考察的是中西方文化差异对虚拟人道德责任判断的影响，主要区别在于自变量和研究对象的不同。本文通过 5 个实验发现，相比于西方文化，中国文化影响下的人们认为虚拟人犯错后需要承担更大的道德责任，这一现象的潜在机制是中国文化(相比于

西方文化)影响下的人们认为虚拟人的心智能力更高(其中情绪感知力比认知自主力具有更强的解释力度)。本文还发现，无论虚拟人的背后主体是真实人类还是人工智能，这种中西方文化差异始终存在。此外，对虚拟人更高的道德责任判断会导致中国文化(相比于西方文化)影响下的人们更倾向于对虚拟人施加道德惩罚。本文通过实证研究将道德责任判断和心智感知的对象拓展到虚拟人上，并揭示了中西方文化差异及其后续影响(道德惩罚)。本文拓展了与文化差异、道德判断、虚拟实体和心智感知相关的文献。

根据您的建议，我们在研究结论与讨论部分“7.2 中西方文化差异对虚拟人道德责任判断的影响”的第二段最后，补充说明了本文和 Sullivan & Wamba (2022)研究的区别和贡献：“与 Sullivan 和 Wamba (2022)的研究不同，本文的研究重点关注中西方文化差异对虚拟人道德责任判断的影响。本文的研究表明，相比于西方文化，中国文化影响下的人们认为虚拟人犯错后需要承担更大的道德责任。本文除了发现感知心智能力的中介外，还探索了虚拟人主体类型的调节作用以及中西方文化差异的后续影响(对虚拟人犯错后的道德惩罚)。”

……

审稿人 3 意见：鉴于作者已经很好的回答了我提出的问题，并且进行了丰富且有效的内容补充，我同意论文的发表。

回应：感谢您的支持，在本轮修改中我们对文章内容进行了进一步的完善。

参考文献

Appel, J., Von Der Pütten, A., Krämer, N. C., & Gratch, J. (2012). Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Advances in Human−Computer Interaction*, *2012*, 1–10.

Aristotle. (1999). *Nicomachean ethics* (T. Irwin. Trans. & Ed.). Cambridge (2nd ed.). Hackett.

Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J. B., Shariff, A., ... Rahwan, I. (2020). Drivers are blamed more than their automated cars when both make mistakes. *Nature Human Behaviour*, *4*(2), 134–143.

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, *181*, 21–34.

Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annual Review of P*sychology, *68*(1), 627−652.

Bryson, J. J., Diamantis, M. E., & Grant, T. D. (2017). Of, for, and by the people: the legal lacuna of synthetic persons. *Artificial Intelligence and Law*, *25*(3), 273–291.

Cameron, C. D., Payne, B. K., & Knobe, J. (2010). Do theories of implicit race bias change moral judgments? *Social Justice Research*, *23*(4), 272–289.

Champagne, M., & Tonkens, R. (2015). Bridging the responsibility gap in automated warfare. *Philosophy & Technology*, *28*(1), 125–137.

Chu, H. D., Li Y. Y., Ye J. H., Hu F. P., He Q., & Zhao L. (2019). Moral judgment about human and robot agents in personal and impersonal dilemmas. *Chinese Journal of Applied Psychology*, *25*(3), 262–271.

[褚华东,李园园,叶君惠,胡凤培,何铨,赵雷. (2019). 个人−非个人道德困境下人对智能机器道德判断研究. *应用心理学*, *25*(3), 262−271.]

Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics*, *26*(4), 2051–2068.

Constantinescu, M., Vică, C., Uszkai, R., & Voinea, C. (2022). Blame it on the AI? On the moral responsibility of Artificial Moral Advisors. *Philosophy & Technology*, *35*(2), 1–26.

Dang, J., & Liu, L. (2022). Implicit theories of the human mind predict competitive and cooperative responses to AI

robots. *Computers in Human Behavior*, *134*, 107300.

Dietze, P., & Knowles, E. D. (2016). Social class and the motivational relevance of other human beings: Evidence from visual attention. *Psychological Science*, *27*(11), 1517–1527.

Dietze, P., & Knowles, E. D. (2021). Social class predicts emotion perception and perspective-taking performance in adults. *Personality and Social Psychology Bulletin*, *47*(1), 42–56.

Eyssel, F. A., & Pfundmair, M. (2015, August). Predictors of psychological anthropomorphization, mind perception, and the fulfillment of social needs: A case study with a zoomorphic robot. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO–MAN)* (pp. 827–832). IEEE, Kobe, Japan.

Fuller, C. J. (2004). *The camphor flame: Popular Hinduism and society in India*. Princeton University Press.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619−619.

Gray, K., Knobe, J., Sheskin, M., Bloom, P., & Barrett, L. F. (2011). More than a body: Mind perception and the nature of objectification. *Journal of Personality and Social Psychology*, *101*(6), 1207–1220.

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, *96*, 505–520.

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125−130.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101−124.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108.

Greenfield, P. M. (2009). Linking social change and developmental change: Shifting pathways of human development. *Developmental Psychology*, *45*(2), 401−418.

Grossmann, I., & Na, J. (2014). Research in culture and psychology: Past lessons and future challenges. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(1), 1−14.

Hakli, R., & Mäkelä, P. (2019). Moral Responsibility of Robots and Hybrid Agents. *The Monist*, *102*(2), 259–275.

Himma, K. E. (2009). Artifcial agency, consciousness, and the criteria for moral agency: What properties must an artifcial agent have to be a moral agent? *Ethics and Information Technology*, *11*(1), 19–29.

Hofstede, G. (1980). Motivation, leadership, and organization: Do American theories apply abroad? *Organizational Dynamics*, *9*(1), 42−63.

Hou, Y. B., & Zhu, Y. (2002). The effect of culture on thinking style of Chinese people. *Acta Psychologica Sinica*, *34*(1), 106–111.

[侯玉波, 朱滢. (2002). 文化对中国人思维方式的影响. *心理学报*, *34*(1), 106–111.]

Huang, Z. H., Jing, Y. M., Yu, F., Gu, R. L., Zhou, X. Y., & Cai, H. J. (2018). Increasing individualism and decreasing collectivism? Cultural and psychological change around the globe. *Advances in Psychological Science*, *26*(11), 2068−2080..

[黄梓航, 敬一鸣, 喻丰, 古若雷, 周欣悦, 张建新, & 蔡华俭. (2018). 个人主义上升, 集体主义式微?——全球文化变迁与民众心理变化. *心理科学进展*, *26*(11), 2068−2080.]

Krumhuber, E. G., Swiderska, A., Tsankova, E., Kamble, S. V., & Kappas, A. (2015). Real or artificial? Intergroup biases in mind perception in a cross-cultural perspective. *PLoS One*, *10*(9), e0137840.

Letheren, K., Kuhn, K.L., Lings, I., & Pope, N. Kl. (2016). Individual difference factors related to anthropomorphic tendency. *European Journal of Marketing*, *50*(5/6), 973–1002.

Liu, B. H., & Peng, K. P. (2012). Challenge and contribution of cultural psychology to empirical legal studies. *Acta Psychologica Sinica*, *44*(3), 413−426.

[刘邦惠, 彭凯平. (2012). 跨文化的实证法学研究：文化心理学的挑战与贡献. *心理学报*, *44*(3), 413−426.]

Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015, March). Sacrifice one for the good of many? People apply different moral norms to human and robot agents. In *2015 10th ACM/IEEE International Conference on Human−Robot Interaction (HRI)* (pp. 117−124). IEEE, Portland, USA: ACM.

McGuire, B.F. (2018). Buddhist uploads. In M. Bass and D.W. Pasulka (Eds.), *Posthumanism: The Future of Homo Sapiens* (pp. 143–153). New York: Macmillan.

Miller, J. G., & Bersoff, D. M. (1992). Culture and moral judgment: How are conflicts between justice and interpersonal responsibilities resolved? *Journal of Personality and Social Psychology*, *62*(4), 541–554.

Morling, B. (2016). Cultural difference, inside and out. *Social and Personality Psychology Compass*, *10*(12), 693−706.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, *67*(6), 949–971.

Na, J., Grossmann, I., Varnum, M. E., Kitayama, S., Gonzalez, R., & Nisbett, R. E. (2010). Cultural differences are not always reducible to individual differences. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(14), 6192−6197.

Ohnuma, R. (2007). *Head, eyes, flesh, and blood: Giving away the body in Indian Buddhist literature*. Columbia University Press.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, *128*(1), 3−72.

Parthemore, J., & Whitby, B. (2014). Moral agency, moral responsibility, and artifacts: What existing artifacts fail to achieve (and why), and why they, nevertheless, can (and do!) make moral claims upon us. *International Journal of Machine Consciousness*, *6*(2), 141–161.

Peng, K. P., Yu, F., & Bai, Y. (2011). Experimental ethics: New challenges and contributions to the understanding of human moral behaviors. *Social Sciences in China*, *182*(6), 15–25.

[彭凯平, 喻丰, 柏阳. (2011). 实验伦理学: 研究、贡献与挑战. *中国社会科学*, *182*(6), 15–25.]

Samuel, G. (1989). The body in Buddhist and Hindu tantra: Some notes. *Religion*, *19*(3), 197–210.

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, *22*(1), 32–70.

Sullivan, Y. W., & Wamba, F. S. (2022). Moral judgments in the age of artificial intelligence. *Journal of Business Ethics*, *178*, 917–943.

Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review*, *96*(3), 506–520.

Wang, E. J., & Yue, G. A. (2006). Moral diversity between eastern and western cultures — a analysis in the perspective of cultural psychology. *Morality and Civilization*, *2*, 52–56.

[王恩界, 乐国安. (2006). 东西方文化背景下的道德观差异——来自于文化心理学视角的分析. *道德与文明*, *2*, 52–56.]

Van Genderen, H. (2018). Do we need new legal personhood in the age of robots and AI? In M. Corrales, M. Fenwick, N. Forgó (eds.), *Robotics, AI and the Future of Law* (pp. 15–55). Springer, Singapore.

Ward, A. F., Olsen, A. S., & Wegner, D. M. (2013). The harm-made mind: Observing victimization augments attribution of minds to vegetative patients, robots, and the dead. *Psychological Science*, *24*(8), 1437–1445.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383−388.

Willard, A. K., & McNamara, R. A. (2019). The minds of god (s) and humans: Differences in mind perception in Fiji and North America. *Cognitive Science*, *43*(1), e12703.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, *18*(7), 600−606.

Young, A. D., & Monroe, A. E. (2019). Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas. *Journal of Experimental Social Psychology, 85,* 103870.

第三轮

审稿人 1 意见：

意见 1：不知道作者自身是否感觉到引言的展开找不到主线，其实想要在这么大的篇幅下使引言逻辑更加清晰，最好的办法是将（不断递进的）多个假设融入到引言每个部分中，明确体现问题是如何推进的。另外，建议实验顺序适当调整，具体如下： →假设 1：文化差异（实验 1、2、4 可改为实验 1a、1b 和实验 1c） →假设 2：中介机制（实验 3 相应改为实验 2，可以去掉调节变量） →假设 3：差异后效（实验 5 相应改为实验 3）。

回应：感谢您的建议。由于与虚拟人相关的实证研究相对较少，在假设推导的过程中，很难直接从文化差异推导到对虚拟人的道德责任判断上。本文的主要贡献在于文化差异到心智能力评价这一环节。因此本文的假设推导逻辑是：先具体阐述文化差异到中介感知心智能力这一环节，并由此推导到对道德责任判断的影响，最后提出研究假设，以保证逻辑上的通顺。本文的两个研究假设难以直接分开融入到引言某个部分中，在本轮修改中，我们还是保留了在引言最后直接呈现两个研究假设的方法。结合意见 2 和意见 3，我们已对引言部分的整体结构进行了调整，对引言部分的内容进行了精简和修改，使引言的逻辑更为清晰。另外，根据您的建议，我们已对本文的实验顺序进行了适当调整，使实验部分的逻辑更加清晰。

意见 2：引言中问题引入似乎有些跳脱、思路略显混乱。修改建议：①基于目前的行文，引言第二段的第一句应提前到第一段最后，并紧接着进一步指出本研究的意义所在。②然后第二段以“其中一个重要问题便是，人们会对虚拟人犯错做出怎样的道德责任判断？”引入本文的研究。③引言第二段列举 3 例单一事件不足以体现“不同文化下的态度不一致”，应补充相应的潜在文化差异证据。④“可见，虚拟人有可能在网络上面进行侵权、抄袭、盗用、泄密等不道德行为，并且不同文化背景下的人们对其态度也不一致。”这一句话较重要，但是前半句是对现状的描述，“可能”宜改为“会”，这是事实、而非推断；后半句应以“可能”做出推断，因为这并非明确的、或已经检验的事实。⑤从句与句之间的关系来看，“那么，相比于真实人类在社交网络上面进行同样的不道德行为，人们又如何看待虚拟人的不道德行为呢？”中的“那么”“又”两个关系词的使用都不合论证逻辑。引言和讨论中还存在多处类似连接词、逻辑词的使用不当问题，请作者认真推敲。

回应：感谢您的建议让我们对引言部分进行了更好的完善。

①我们已将引言第二段的第一句“然而，在社会关注和资本加持的驱动下，相关的法律和道德伦理问题也会不断产生。”提前到引言第一段的最后。另外，我们在第二段的最后指出了本研究的意义所在：“本文将通过五项实验考察中西方文化差异对虚拟人道德责任判断的影响和机制，对虚拟实体道德判断的相关研究进行了延伸，丰富了文化差异、心智感知等相关文献，还为虚拟人的设计、运营和在道德伦理上的治理问题提供了一定的实践指导意义”。

②我们已将引言第二段的第一句修改为：“其中一个重要问题便是，人们会对虚拟人犯错做出怎样的道德责任判断？”

③根据您的建议，我们额外补充了虚拟人在西方社交媒体经常散布谣言和诋毁选举人的实例，这与国内媒体实名责任制的管控不同，体现了潜在的中西方文化差异。具体来说，我们在引言第二段的第 6-9 行增加了一句：“研究发现，在欧美等国外的政治大选中也存在社

交虚拟人在线诋毁候选人的情况(Kollanyi et al., 2016)，他们也会传播未经证实的健康声明，甚至对公众健康造成危害(Allem et al., 2020)，但并未得到治理和管控，如若明确表明机器人身份，他们依然可以畅所欲言。”

④我们已将引言第二段的第 11-12 行修改为：“可见，虚拟人会在网络上面进行侵权、抄袭、盗用、诋毁等不道德行为，并且不同文化背景下的人们可能对其态度也不一致。”

⑤我们已将引言第二段的第 12-13 行修改为：“相比于真实人类在社交网络上面进行同样的不道德行为，人们会如何看待虚拟人的不道德行为呢？”。根据您的建议，我们对引言和讨论中类似连接词、逻辑词使用不当的问题进行了修改。

意见 3：引言整体结构，建议结合上述意见 1 再调整下。此外：①《心理学报》引言有字数限制，尤其是对于本文这么大的篇幅来看，引言更应是寸土寸金。审稿人指出应加深文化理论方面的阐述，这符合《学报》重理论意义的宗旨，但作者不宜在“1.1 中西方文化差异对道德判断的影响”几乎罗列式地、完全脱离道德判断地去介绍文化理论。②“虚拟人”是本研究的核心研究对象，从目前的结构来看，“1.2 虚拟人的定义及研究范畴”要么提前到引言第二部分（即：1.1 虚拟人的定义及研究范畴），要么压缩后置于“对虚拟实体的道德责任判断”部分的第一段，否则很突兀。

回应：感谢您的建议。

①我们已对引言部分的内容进行了精简和修改，使引言的字数更贴近《心理学报》的要求。另外，我们修改了“1.1 中西方文化差异对道德判断的影响”，具体内容如下所示：

1.1 中西方文化差异对道德判断的影响

文化会对人们的心理与行为产生重要影响(侯玉波, 朱滢, 2002)。在文化心理学中，文化泛指社会成员之间所共有的价值观、规范、思维方式和行为方式等(Hofstede, 1980; Morling, 2016; Na et al., 2010; 黄梓航 等, 2018)。在文化心理学领域，许多关于东西方文化差异的研究都是基于个人主义/集体主义的文化理论(Triandis, 1989)。

先前的研究发现，东西方文化差异会影响人们的道德判断。西方文化强调以个人自由、权利、公正、关爱和宽恕等为导向的道德观，而东方文化强调以集体和个人责任为导向的道德观(彭凯平等, 2011; 王恩界, 乐国安, 2006)。Miller 和 Bersoff (1992)的研究指出，西方人更强调公正伦理，而东方人更强调责任伦理。个人主义/集体主义文化也会影响人们的道德决策，因为它们涉及到个人利益还是集体利益优先的信念(Oyserman et al., 2002)。在本文中，基于个人主义/集体主义的文化心理学理论，我们关注到了中西方文化差异对虚拟人道德责任判断的影响。与个人主义文化(如美国)相比，集体主义文化(如中国)下的人们会有更高的换位思考能力(Wu & Keysar, 2007)和拟人化倾向(Letheren et al., 2016)。相比于西方文化，中国文化影响下的人们是否会对虚拟人的心智能力评价更高，进而认为虚拟人应该为不道德行为承担更大的道德责任？这是本文想要回答的问题。

②我们已对原文中“1.2 虚拟人的定义及研究范畴”的内容进行了精简，并放到了“1.2 对虚拟人的道德责任判断”的第一段，使引言部分的整体结构更为合理。

意见 4：实验部分的一些细节问题。①与前文实验相同的内容直接一句话带过即可，一些实验程序描述也无需重复说明，显得特别冗余。如“2.2 实验流程”，其中第一句和“2.1 实验设计与样本的最后一句”完全重复了，第二句的前半句和后半句也几乎是重复的。②正文全部采用“A 和 B 的交互作用对 C 有显著影响”这样的表述，方差分析中交互作用是这样汇报的吗？是不是改成“A 和 B 影响 C 的交互作用显著”更合适？③本文条形图的图注全是交互作用，但这些条形图全都看不出交互作用显著，至多能体现简单效应。建议修改为类似于描述性统计的表述，参照许丽颖等（2022）一文。④实验 2 和实验 4 的操纵检验中采用的是

回归分析，请具体汇报进行了何种类型的回归分析。⑤作者能够采用平均值的置信区间做差异检验值得肯定，但是对于一些非核心的效应只报告方差分析结果可能更好（F 值、p 值、效应量、置信区间），没必要报告单组的平均值、标准差、置信区间，否则显得太乱了。⑥结合意见 1，从本研究的旨趣来看，既然对真人组道德判断的文化差异不显著，那么实验 3 就没有必要做有调节的中介了，只需要做“文化→心智感知→道德判断”的简单中介即可，这样会使结果汇报更加清晰。现在这样汇报当然也没大问题，仅供作者参考。⑦附录二中，文化类型的主效应 $p = 0.049 < 0.05$，怎么是“边际显著”呢？

回应：感谢您的建议。

①根据您的建议，我们已删去全文实验流程中存在重复和冗余的内容。

②结合您的建议和心理学报已发表的文章描述，我们将交互效应结果的表述全文统一修改为“A 和 B 对 C 的交互效应显著”。

③参考许丽颖等(2022)一文，我们已全文修改条形图的图注。例如，将“文化类型与博主类型的交互作用对道德责任判断的交互影响”改为“不同文化类型下对虚拟人和真人的道德责任判断评分”。

④我们已在实验 2 和实验 4 的操纵检验部分具体汇报了采用的回归分析类型。

⑤根据您的建议，我们已全文删除了非核心效应结果中单组的平均值、标准差和置信区间等数据，使数据结果部分的呈现更为清晰。

⑥感谢您的建议，为了实验的完整性和汇报数据的透明公开性，我们还是决定保留中介实验中的真人组。此外，结合第二位审稿人的意见，我们在复制中介效应的实验中仅针对虚拟人组进行复制。我们采用单因素的实验设计，并按照您所提到的简单中介效应汇报实验结果，具体内容见正文部分的附录三。

⑦我们已修改为“显著”的表述，并全文检查了是否存在类似的问题。

.....

审稿人 2 意见：

作者根据的评审专家的意见对文献综述、假设提出的理论论述和逻辑进行了较大幅度的提升，论述更好地围绕着中西文化差异展开，修订了其中论述不清或有误的部分。总的来说，较好地回应了评审专家就理论层面提出的意见。但是对实验方面提出的意见，并没有做出实质性的回应和增补。特别是，意见：“全文 5 个实验虽然采用了不同的不道德行为，但是情境的设定都是一样的，都选择了女性的博主。这样单一的情境设定会也会在一定程度上影响结果的可推广性。而商业实践中的虚拟人类型是非常丰富多样的，如游戏人物、品牌代言人，虚拟偶像等等。建议采用其他的实验情境设定来 replicate 其中的部分实验。”尽管作者认为实验 5 的设定为偶像而非博主。但是采用的实验材料相似度仍较大。为了增强实验结果的可靠性，建议作者采用其他具有一定跨度和差异的实验情境和实验材料 replicate 其中的 1 个或部分实验。

回应：感谢您的支持和建议。根据您的建议，我们采用男性实验材料复制了中介效应实验（正文实验 2）。结合第一位审稿人的意见，在已经验证真人组不存在中西方文化差异后，为了使汇报的结果更加清晰直观，我们在复制实验中仅聚焦于虚拟人组，不再探讨真人组的中西方文化差异。考虑到正文的内容篇幅，我们将复制实验的内容放在附录三中，并在正文“7.4 研究局限与未来研究方向”第五段的第 7-9 行提到“另外，社交网络上虽然女性虚拟人占比更大，但是也存在许多男性虚拟人。本研究采用男性虚拟人的实验材料复制了中介实验(见附录三)，并得出了一致的结果”。

参考文献

Allem, J. P., Escobedo, P., & Dharmapuri, L. (2020). Cannabis surveillance with Twitter data: Emerging topics and social bots. *American Journal of Public Health*, *110*(3), 357–362.

Kollanyi, B., Howard, P. N., & Woolley, S. C. (2016). Bots and automation over Twitter during the first US presidential debate. *Comprop Data Memo*, *1*, 1−4.

Cameron, C. D., Payne, B. K., & Knobe, J. (2010). Do theories of implicit race bias change moral judgments? *Social Justice Research*, *23*(4), 272–289.

Gray, K., Knobe, J., Sheskin, M., Bloom, P., & Barrett, L. F. (2011). More than a body: Mind perception and the nature of objectification. *Journal of Personality and Social Psychology*, *101*(6), 1207–1220.

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*(1), 1–22.

Hofstede, G. (1980). Motivation, leadership, and organization: Do American theories apply abroad? *Organizational Dynamics*, *9*(1), 42−63.

Hou, Y. B., & Zhu, Y. (2002). The effect of culture on thinking style of Chinese people. *Acta Psychologica Sinica*, *34*(1), 106–111.

[侯玉波, 朱滢. (2002). 文化对中国人思维方式的影响. *心理学报*, *34*(1), 106–111.]

Huang, Z. H., Jing, Y. M., Yu, F., Gu, R. L., Zhou, X. Y., & Cai, H. J. (2018). Increasing individualism and decreasing collectivism? Cultural and psychological change around the globe. *Advances in Psychological Science*, *26*(11), 2068−2080..

[黄梓航, 敬一鸣, 喻丰, 古若雷, 周欣悦, 张建新, & 蔡华俭. (2018). 个人主义上升, 集体主义式微?——全球文化变迁与民众心理变化. *心理科学进展*, *26*(11), 2068−2080.]

Letheren, K., Kuhn, K.L., Lings, I., & Pope, N. Kl. (2016). Individual difference factors related to anthropomorphic tendency. *European Journal of Marketing*, *50*(5/6), 973–1002.

Miller, J. G., & Bersoff, D. M. (1992). Culture and moral judgment: How are conflicts between justice and interpersonal responsibilities resolved? *Journal of Personality and Social Psychology*, *62*(4), 541–554.

Morling, B. (2016). Cultural difference, inside and out. *Social and Personality Psychology Compass*, *10*(12), 693−706.

Na, J., Grossmann, I., Varnum, M. E., Kitayama, S., Gonzalez, R., & Nisbett, R. E. (2010). Cultural differences are not always reducible to individual differences. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(14), 6192−6197.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, *128*(1), 3−72.

Peng, K. P., Yu, F., & Bai, Y. (2011). Experimental ethics: New challenges and contributions to the understanding of human moral behaviors. *Social Sciences in China*, *182*(6), 15–25.

[彭凯平, 喻丰, 柏阳. (2011). 实验伦理学: 研究、贡献与挑战. *中国社会科学*, *182*(6), 15–25.]

Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review*, *96*(3), 506–520.

Wang, E. J., & Yue, G. A. (2006). Moral diversity between eastern and western cultures — a analysis in the perspective of cultural psychology. *Morality and Civilization*, *2*, 52–56.

[王恩界, 乐国安. (2006). 东西方文化背景下的道德观差异——来自于文化心理学视角的分析. *道德与文明*, *2*, 52–56.]

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, *18*(7), 600−606.

Xu, L. Y., Yu, F., & Peng, K. P. (2022). Algorithmic discrimination causes less desire for moral punishment than human discrimination. *Acta Psychologica Sinica*, *54*(9), 1076−1092.

[许丽颖, 喻丰, 彭凯平. (2022). 算法歧视比人类歧视引起更少道德惩罚欲, *心理学报*, *54*(9), 1076−1092.]

第四轮

审稿人 2 意见：

作者对评审意见做出了较好地回应了，达到了发表的水平，建议录用。

编委意见：

The paper looks great. Please accept.

An excellent paper.

主编意见：

论文达到学报发表要求。同意发表。只是有个小问题请作者斟酌：5 项实验，5 个实验，哪种表达更恰当？我个人感觉，说“完成 5 项实验研究”没问题，但说“完成 5 项实验”似乎就不如说“完成 5 个实验”。