

化身机器人对个体道德两难决策的影响

黄欣语, 麻珂*

西南大学心理学部, 重庆

收稿日期: 2026年9月15日; 录用日期: 2026年10月12日; 发布日期: 2026年10月21日

摘要

具身认知理论认为, 认知并非独立于身体和环境而发生, 而是在认知、身体与环境的交互中形成。沉浸式虚拟现实(IVR)为探讨身体拥有感错觉如何影响认知提供了重要技术手段。本研究采用单因素被试内设计, 通过同步或异步的视觉-动觉多感官刺激操纵被试的身体拥有感。在同步条件下, 被试可以自由操控机器人化身, 并对机器人产生身体拥有感, 从而考察化身成为标准人形机器人后对个体道德两难决策的影响。结果发现, 与异步条件相比, 同步条件下个体对不作为的普遍偏好(I参数)显著减弱, 而结果敏感性(C参数)和规范敏感性(N参数)未受到显著影响。这表明, 对标准人形机器人化身产生身体拥有感错觉会促使人们在道德困境中更倾向于采取行动, 而非选择不作为。本研究为远程机器人应用与人机交互设计中的道德决策考量提供了实证参考, 未来研究可进一步扩大样本, 以验证其作用机制。

关键词

具身认知, 身体拥有感错觉, 道德决策, 机器人化身

The Influence of Embodied Robot Avatars on Individuals' Moral Dilemma Decision-Making

Xinyu Huang, Ke Ma*

School of Psychology, Southwest University, Chongqing

Received: September 15, 2026; accepted: October 12, 2026; published: October 21, 2026

Abstract

Embodied cognition theory posits that cognition is not independent of the body and environment,

*通讯作者。

文章引用: 黄欣语, 麻珂(2026). 化身机器人对个体道德两难决策的影响. *心理学进展*, 16(10), 232-243.
DOI: 10.12677/ap.2026.1610488

but is formed in the interaction among cognition, body, and environment. Immersive virtual reality (IVR) provides a technical means to explore the influence of the body ownership illusion (BOI) on cognition. Using a single-factor within-subjects design (synchronous vs. asynchronous conditions), the present study examined the influence of embodied standard humanoid robot avatars on individuals' moral dilemma decision-making. In the synchronous condition, participants could freely control the robot avatar and develop a sense of body ownership over it, thereby examining the influence of embodying a standard humanoid robot on individuals' moral dilemma decision-making. Results showed that the general preference for inaction (I parameter) was significantly reduced in the synchronous condition relative to the asynchronous condition, whereas sensitivity to consequences (C parameter) and sensitivity to moral norms (N parameter) did not differ significantly between conditions. These findings suggest that embodying a standard humanoid robot avatar can weaken individuals' general preference for inaction in moral dilemmas, providing empirical implications for moral decision-making in teleoperated robot applications and human-robot interaction design. Future research with larger samples is needed to further examine the underlying mechanisms.

Keywords

Embodied Cognition, Body Ownership Illusion, Moral Decision-Making, Robot Avatar

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

具身认知理论认为, 思维植根于感觉运动经验的具体细节, 认知、身体与环境相互交织而非彼此割裂。这一观点挑战了传统认知科学中“心理表征不受身体影响”的主张(Foglia & Wilson, 2013)。近年来, 具身认知日益受到研究者的关注与检验。依循这一观点, 当个体拥有一个在结构和属性上不同于原初身体的躯体时, 其认知也会随之改变。沉浸式虚拟现实(Immersive Virtual Reality)的发展为验证这一设想提供了切实可行的技术路径: 实验者借助头戴式显示器(HMD)、惯性系统与红外摄像头捕捉被试动作, 并据此驱动虚拟化身(avatar)运动, 使被试获得中介化的具身体验, 进而产生身体拥有感错觉。身体拥有感错觉(Body Ownership Illusion)指个体将某一客体误认作自身身体一部分的错觉(Kilteni et al., 2015)。经典橡胶手错觉(RHI)实验便是有力例证: 主试将被试真实手隐藏于视线之外, 把橡胶手置于其面前、与真实手位置大致相当, 随后同时用刷子刷拂真实手与橡胶手, 结果多数被试报告橡胶手仿佛属于自己身体(Botvinick & Cohen, 1998); 当橡胶手受到威胁时, 被试还会产生相应的皮层焦虑反应(Ehrsson et al., 2007)。此后, 研究者进一步尝试通过实验操纵诱发“灵魂出窍”式的全身拥有感体验, 从人体模型到虚拟化身(Petkova et al., 2011), 并利用同步/异步、第一/第三视角、有镜/无镜等操纵较稳定地建立全身拥有感错觉, 进而考察其对个体认知与行为的影响(Petkova et al., 2011; Kilteni et al., 2015)。

1.1. 化身机器人对个体的影响

随着人工智能技术不断取得突破, 远程机器人有望深度融入人类生活的各个领域, 使人们能够在遥远之处“附身”机器人, 并与当地的人进行互动。然而, 已有研究多聚焦于如何设计高度人性化的化身, 较少探讨拥有机器人化身对个体的影响(Nishio et al., 2012; Aymerich-Franch et al., 2016)。比如, 有研究发现, 操控远程机器人的个体会对该机器人产生身体拥有感错觉(Alimardani et al., 2013), 而这一错觉如何

影响个体的认知、行为与选择,以进行的研究发现相对较少。目前,研究者就拥有机器人化身在个体道德、情感与社会态度等方面的影响展开初步探索。例如,Aymerich-Franch 等人(2020)发现,即使机器人化身的辱骂行为并非个体自身所为,被试也会感到内疚与羞耻,并倾向于道歉。也有研究指出,机器人化身有助于提高人们对机器人的接受度(Farizon et al., 2022); Farizon 等人(2022)进一步发现,与仅通过视觉刺激诱发的具身感相比,由动作同步诱发的主动控制感(sense of agency)才是改善人们对机器人态度的关键。值得注意的是,匿名操控远程机器人化身还可能诱发低道德行为(Shaheen et al., 2025)。

1.2. 道德两难决策

总之,前人发现似乎表明,拥有机器人化身的具身错觉能够影响人们的情绪反应与社会偏好,但该效应能否作用于更为复杂的道德决策过程呢?这也就是本文聚焦的研究目的:化身机器人是否会影响人们的道德两难决策?尤其是考虑到,如果未来远程机器人大规模进入生活,当人们通过远程机器人做出决策时,他们对远程机器人产生的身体拥有感错觉,认为“我就是这个远程机器人”的感觉,是否会改变他们的道德选择。

经典的非个人困境电车难题设想了这样的情境:一辆失控的电车正驶向五个人,你可以在控制室通过按动按钮改变电车轨道,使其转向另一条只有一个人的轨道;你也可以选择不按按钮,让五个人死亡(foot, 1967)。随后,Thomson (1976)提出了经典的个人困境天桥难题:你站在天桥上,看到一辆失控的电车驶向五个人,你可以把桥上的一个胖子推下去阻挡电车,从而只牺牲一人;你也可以选择不推,让五个人死亡。天桥困境属于个人困境,因为它涉及直接的身体接触或情感联结;电车难题则以间接、非接触的方式使他人受害,属于非个人困境。无论是个人困境还是非个人困境,功利论都倾向于牺牲少数以拯救多数,以实现整体利益的最大化与成本的最小化(Mill, 1992);而道义论则强调权利与义务优先于利益权衡,坚持无论何种情况都不应主动做出伤害他人的行为(康德, 1957)。前人研究认为,在非个人困境中,人们往往倾向于功利论的选择,而在个人困境中,人们则会做出道义性的决策。Greene 等人(2001, 2008)针对这两种不同的选择倾向提出了双过程理论(process dissociation analysis),重点强调了自动情感过程在道德判断中的作用,反驳了早期理性主义认为道德判断是理性思考结果的观点。该理论认为,功利论的选择更多依赖认知控制,需要克服自动化的情绪反应;道义论的选择则更多由自动情感过程驱动。由于个人困境会引发更强烈的情绪反应,人们在个人困境中更倾向于道义论。双过程理论得到了众多研究的支持与补充(Koenigs et al., 2007)。

然而,以电车难题为代表的传统困境范式也受到了一些研究者的质疑,其中突出的两点是:情境不现实、生态效度有限(Bauman et al., 2014),以及范式未能系统操纵行为结果与道德规范、难以区分决策背后的不同心理过程(Gawronski et al., 2017)。以天桥困境为例,被试选择推下胖子可能出于三种原因:一是对道德结果敏感性较高,二是对道德规范敏感性较低,这两种原因确实可以反映功利论倾向的决策,但个体做出功利论倾向决策背后的具体原因是传统困境范式无法区分的;此外还有第三种可能,即个体只是偏好采取行动本身,而不考虑推下胖子能否阻止电车、挽救五人的生命,这种原因不能被定义为功利论倾向。为更准确地区分、定义和量化个体的道德决策倾向,Gawronski 等人(2017)提出了 CNI (consequence, norm, inaction)模型,将传统道德困境按两个维度(行为结果为收益大于或者小于成本,道德规则为禁止或者倡导)分为四类共 24 个困境:(1) 禁止性规范禁止行动,而行动对整体幸福的好处大于行动的成本;(2) 禁止性规范禁止行动,而行动对整体幸福的收益小于行动的成本;(3) 规定性规范规定行动,而行动对整体幸福的收益大于行动的成本;(4) 规定性规范规定行动,而行动对整体幸福的收益小于行动的成本;并通过多项式建模估计三个参数:对行为结果的敏感性(C)、对道德规范的敏感性(N)以及一般性不作为偏好(I)。之后,Körner 等人(2020)为弥补 CNI 模型在个体差异测量上的局限,将困境扩充为 48 个。

目前 CNI 模型已被广泛用于道德决策研究(Lu et al., 2022), 围绕其效度的质疑与回应也推动了该模型的完善(Gawronski et al., 2020), 本研究采用了 CNI 模型考察化身机器人对个体道德决策的影响。

1.3. 化身与道德两难决策

大量具身认知研究表明, 虚拟化身体验能够显著改变人们的自我感知、社会态度与实际行为。人们在沉浸式 VR 环境中具身拥有不同身份的虚拟身体后, 虚拟化身带来的具身重塑效应具有普遍性, 能够在一定程度上重塑人们的自我表征与心理加工方式(Yee & Bailenson, 2007), 从而引起相应的认知与情绪偏移。例如, 在虚拟环境中, 化身成为爱因斯坦可提升被试的认知任务表现(Banakou et al., 2018)。基于化身效应的广泛影响, 近年来研究者逐渐将具身视角引入道德心理领域, 探讨虚拟身体体验对个体道德认知与道德行为的作用。相关研究发现, 在沉浸式虚拟环境中以化身实施攻击性行为会提升个体的攻击性感受与行为(Persky & Blascovich, 2008); 对机器人的拟人化与对人的去人性化之间也存在密切关联(Dang & Liu, 2024)。

与此同时, 虚拟现实技术也为传统道德两难范式提供了生态化改进路径: 沉浸式道德情境能够改变个体的决策过程与倾向, 并支持结合皮电、心率等生理指标考察道德认知(Pucci & Gagliardi, 2025; Salagean et al., 2024)。然而, 现有研究仍存在明显争议与研究空白。当前 VR 道德研究多聚焦于沉浸式情境体验本身, 较少剥离并单独探讨虚拟化身属性对道德两难决策的特异性影响。少数专门探究化身效应的研究也未得出稳定结论: 部分研究发现个性化化身仅提升道德决策中的生理唤醒, 却无法改变最终决策结果(Salagean et al., 2024); 另有研究发现对上帝化身的具身拥有对个体道德两难决策无显著作用(Frisanco et al., 2024)。总体而言, 现有研究尚未发现稳定的化身效应, 这一结果可能源于两方面局限: 其一, 以往研究多采用争议较大的经典道德两难范式, 测量灵敏度不足, 难以捕捉细微的道德加工差异; 其二, 已有研究多选用人类类别的化身, 缺乏对非人类、低情绪性化身的探讨。

为弥补上述不足, 本研究选取机器人化身作为实验材料。机器人通常被人们判断为缺乏情绪、脱离人类道德主体的非人类客体, 而个体的情绪加工是道德规范敏感性的重要基础(Gawronski et al., 2017; Körner et al., 2020)。据此可推测, 当个体通过身体拥有感错觉与机器人化身完成身份同化时, 其自动情绪反应可能受到抑制, 进而降低对道德规范的感知敏感性, 表现出更强的功利主义决策倾向。此外, 人机道德比较研究进一步为本研究提供了理论支撑: 研究者已系统考察了人们对人类与机器人在经典道德困境中决策的差异评价, 发现判断类型与困境结构等因素会调节人们对两类主体的道德评价(Scheutz & Malle, 2021; Malle et al., 2025)。基于这一认知偏差, 被试在化身为机器人后, 其一般性不作为偏好可能会发生改变。因此, 本研究依托 CNI 模型, 从道德规范敏感性、功利结果倾向与不作为偏好三个维度提出如下假设: H1: 机器人化身诱发的具身体验会降低个体的道德规范敏感性(N), 使其更倾向于做出功利主义决策; H2: 化身机器人会削弱个体在道德困境中的一般性不作为偏好(I)。为检验上述假设, 本研究采用单因素被试内实验设计, 自变量为是否对机器人化身产生身体拥有感错觉, 通过视觉-动觉的同步性操纵分为同步与异步两种实验条件。

2. 材料与方法

2.1. 被试

我们使用 G*Power 3 (Faul et al., 2007)进行先验样本量估计, 设定中等效应量($d = 0.5$)、统计检验力($1 - \beta$) = 0.80、显著性水平 $\alpha = 0.05$ (双尾), 计算得到所需样本量为 34。共招募 40 名被试, 其中 2 名被试因未认真作答被剔除, 最终保留 38 名被试(18 名女性, 20 名男性), 年龄 18~24 岁($M = 20.47$, $SD = 1.81$)。被试均通过大学校园广告招募, 职业均为学生, 认真完成实验的被试获得了 60~70 元人民币的报酬。整个实验过程中被试对实验目的不知情。本研究经西南大学心理学部伦理委员会审查批准。

2.2. 实验设备

本实验在虚拟现实实验室中进行, 使用 Vizard 6.0 平台搭建标准化虚拟实验环境, 并结合 HTC Vive 虚拟现实头盔、定位器、传感器与手柄实现被试对机器人化身身体拥有感错觉的操纵。系统使用 Vizard 6.0 自带的标准人形机器人模型。本实验采用 Vizard 6.0 内置标准人形机器人模型。该模型为白色金属质感, 身高与参与者大致匹配, 类人身体比例, 头部为简化面部设计, 无明确面部表情, 无性别化服饰与毛发, 四肢为分段式机械关节, 整体呈现中性、非情绪化、低至中等拟人化的机器人外观。选择该模型是因为本研究关注非人类、低情绪性化身的具身效应。需要说明的是, 机器人外观可能影响拟人化、心智感知与道德期待, 因此本研究结果应限于此类标准人形机器人, 不能直接推广到高拟人化、类人皮肤、非人形或高度情绪化的机器人化身。在同步条件下, 被试以第一人称视角进入虚拟场景, 低头可直接观察到对应自身姿态的机器人躯体; 机器人的头部、双臂与双腿均可实时跟随被试的身体运动, 场景正前方设置虚拟镜面, 使被试能够完整观察自身机器人化身的整体样貌和动作(见图 1(a))。系统实时捕捉被试头部、四肢的运动信息, 并驱动虚拟机器人化身产生完全同步的肢体动作, 从而通过视觉动觉信息的同步匹配来有效诱发身体拥有感错觉(见图 2)。已有研究表明, 第一人称视角是全身身体拥有感错觉产生的关键条件(Petkova et al., 2011), 因此本研究通过视角与同步性的结合强化具身体验。在异步条件下, 实验将观察视角切换为第三人称视角(见图 1(b)), 同时该条件下的机器人化身执行预设的重复动作, 其运动与被试的真实动作完全无关, 以此有效排除身体拥有感错觉的产生(Aymerich-Franch et al., 2020)。

2.3. 程序

被试到达实验室后, 主试为其佩戴 VR 头盔与传感器, 正式进入虚拟空间。首先, 被试根据标准化指令依次活动头部与四肢, 完成约一分钟的基础设备适配与动作校准; 随后进入四分钟自由活动阶段, 为充分诱发身体拥有感错觉, 主试提示被试尽量保持肢体活动、避免长时间静止。整体虚拟化身适应过程共计五分钟。其中, 同步条件下的机器人化身可实时匹配被试的肢体运动; 异步条件下的机器人化身则持续执行固定重复动作, 与被试自身运动完全无关。为平衡顺序效应与练习效应, 本研究对两种实验条件进行被试间顺序平衡: 半数被试首次实验完成同步条件任务、间隔一周后再次来到实验室完成异步条件任务, 其余被试以相反顺序完成实验。单次化身操纵结束后, 被试首先填写具身拥有感问卷以检验操纵有效性, 随后完成道德两难决策任务; 所有被试间隔一周再次返回实验室, 完成另一实验条件的全部实验流程。

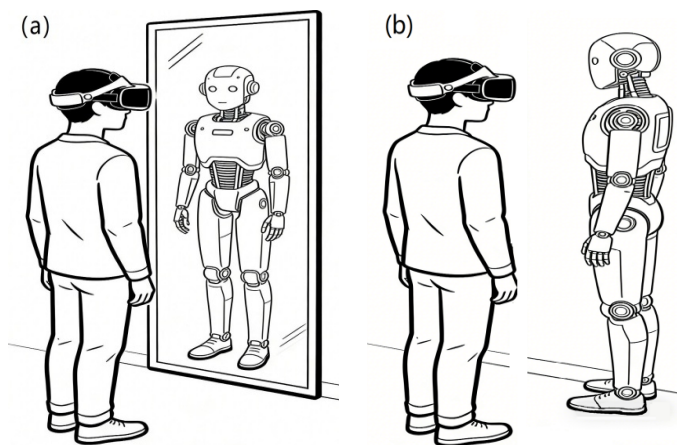


Figure 1. (a) In the 1pp condition, a participant saw the robot through the mirror in the virtual environment. (b) In the 3pp condition, a participant saw the robot standing along with her/him

图 1. (a) 在第一人称视角条件下, 参与者通过虚拟环境中的镜子看到机器人。 (b) 在第三人称视角条件下, 参与者看到机器人和自己一起站着

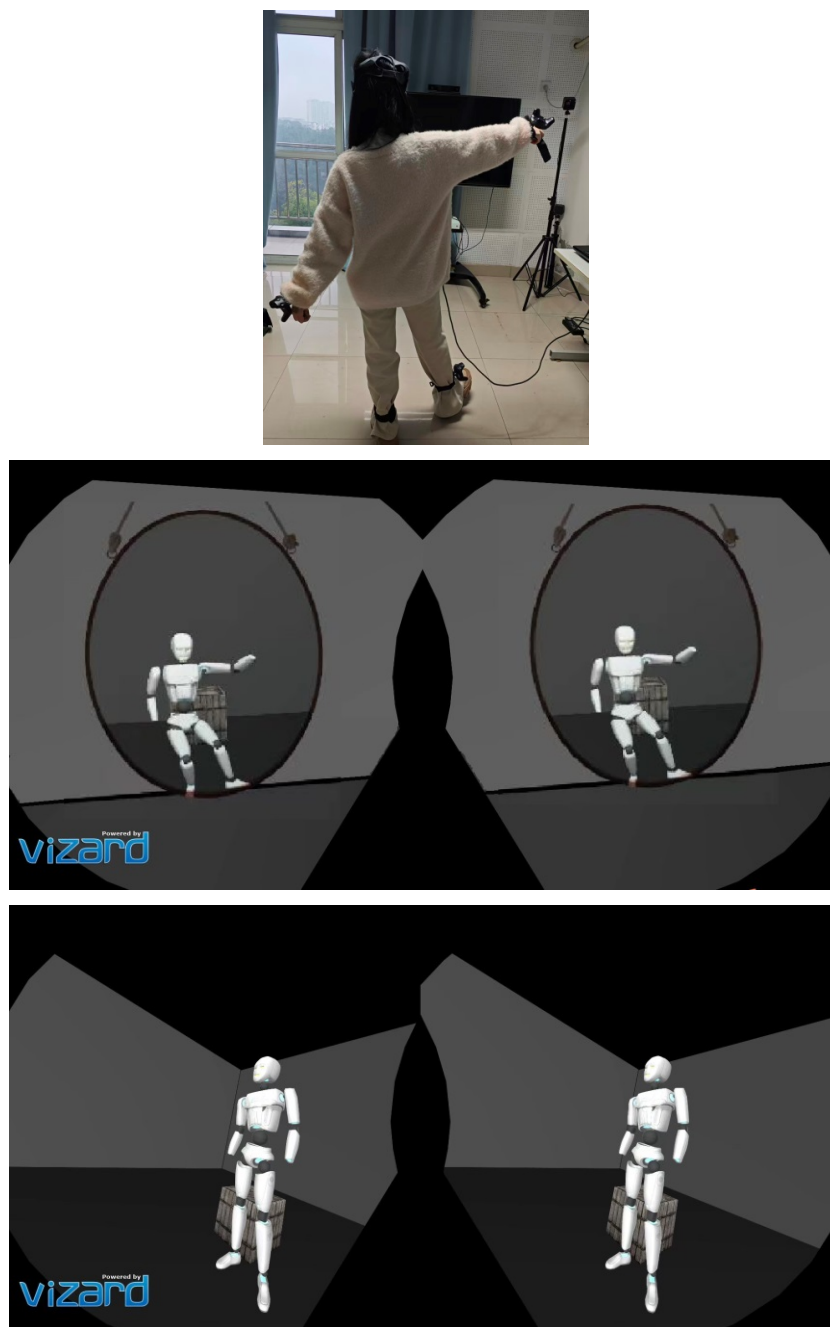


Figure 2. A participant wore the VR HMD on her head and sensors on her wrists and ankles, as she made a specific movement (above). In the 1pp condition, a participant saw the robot in a mirror making the same movement as her (middle). In the 3pp condition, a participant saw the robot standing along with her, making an unrelated movement (below)

图 2. 一位参与者头戴 VR 头显, 手腕和脚踝上带有传感器, 她正在做特定动作(上图)。在第一人称视角条件下, 参与者通过镜子看到机器人做与她相同的动作(中图)。在第三人称视角条件下, 参与者看到机器人站在她旁边, 做与她无关的动作(下图)

2.4. 因变量

2.4.1. 具身拥有感

我们通过 5 个题目测量被试的具身拥有感, 题目条目参照以往机器人化身具身研究的问卷设置

(Aymerich-Franch et al., 2016, 2020), 采用-3 到+3 的 7 点计分(0 表示中立回答, -3 表示强烈不同意, +3 表示强烈同意)。问卷包括身体拥有感问题(mylimb、mybody、mirror)、控制感问题(agency)以及控制问题(twobodies)等题目; 其中, 身体拥有感题目用于考察被试对机器人化身的身体所有权错觉(Botvinick & Cohen, 1998; Kilteni et al., 2015), mirror 题目考察被试在镜像中是否将机器人身体知觉为自身身体(Petkova et al., 2011), 控制感题目考察被试对化身运动的主动控制感(Farizon et al., 2022); 控制题用于考察被试是否体验到“拥有两个身体”(Aymerich-Franch et al., 2016), 其作答水平在两种条件下应保持一致。本研究中量表的内部一致性信度为 $\alpha = 0.938$ 。

2.4.2. 道德决策

本研究采用 Körner 等人(2020)基于 CNI 模型编制的可用于个体差异研究的 48 题道德困境题库, 选取其中 44 个正式实验材料。这是由于溯因困境(abduction dilemma)被证实构念效度较低、无法稳定测量道德决策倾向(Körner et al., 2020), 本研究予以剔除。所有困境题目以固定随机顺序呈现, 被试需要判断情境中的行为是否可接受并做出二择一决策。每次决策完成后, 被试需采用 1~5 点量表分别评价本次决策的主观难度与情绪卷入程度。为缓解长时间任务带来的疲劳效应与注意力衰减, 实验中途设置 1 分钟强制休息时间, 休息结束后由被试自主开启后续实验流程。实验末尾设置注意力检测题目: 以嵌入式填充题的形式呈现, 要求被试忽略当前问题与所有选项、直接按压空格键进入下一页, 以此筛选未认真阅读指导语、随意作答的无效被试数据。本研究通过 E-Prime 2.0 软件全程记录被试的道德决策行为数据。

3. 分析与结果

3.1. 具身拥有感

为验证同步、异步两种实验条件下被试的具身体验差异, 本研究对具身拥有感问卷数据进行统计分析。由于数据不满足参数检验前提, 本研究以实验条件(同步/异步)为被试内变量, 对问卷各题目得分进行 Wilcoxon 符号秩检验。结果显示, 具身拥有感全部五道题目在两种条件下均存在显著差异($p < 0.001$), 同步条件的具身拥有感得分显著高于异步条件。描述性统计结果表明, 同步条件下问卷各题得分中位数介于 1~2 之间, 其中控制题得分中位数最低(中位数 = 1); 异步条件下多数题目得分中位数为-2, 控制题得分中位数为-1.63。上述结果证实, 同步条件成功诱发了被试对机器人化身的拥有感错觉, 而异步条件未产生显著具身体验, 本研究的实验操纵有效, 结果见表 1。

Table 1. Descriptive statistics and Wilcoxon signed-rank test results for each item of the body ownership questionnaire (N = 38)
表 1. 具身拥有感问卷各题描述统计与 Wilcoxon 符号秩检验结果(N = 38)

题目	组别	平均值	标准差	最小值	最大值	中位数	Wilcoxon 符号秩检验(p)
mylimb	同步	1.5845	1.14786	-2	3	2	<0.001
mylimb	异步	-1.4155	1.44026	-3	2	-2	<0.001
mirror	同步	1.3629	1.33869	-2	3	1.6974	<0.001
mirror	异步	-1.5540	1.56732	-3	2	-2	<0.001
control	同步	0.4488	1.60153	-3	2	1	<0.001
control	异步	-1.3560	1.50964	-3	2	-1.6316	<0.001
agency	同步	2.0000	0.77110	0	3	2	<0.001

续表

agency	异步	-1.5263	1.51545	-3	2	-2	<0.001
mybody	同步	1.4155	1.28137	-2	3	2	<0.001
mybody	异步	-1.3352	1.75532	-3	2	-2	<0.001

3.2. 道德决策

本研究参照 Körner 等人(2020)的道德决策数据分析范式，对所有被试的道德两难任务数据进行整理。统计过程中汇总每名被试的 8 项行为指标，包含 4 项行为指数与 4 项不作为指数，对应 11 类道德困境情境，最终各项指标得分区间为 0~11 分。随后采用 multiTree 软件(Moshagen, 2010)拟合 CNI 道德决策模型，分别估计被试在同步、异步条件下的结果敏感性(C 参数)、道德规范敏感性(N 参数)与一般性不作为偏好(I 参数)(Gawronski et al., 2017)。

本研究采用配对样本 t 检验对比两种实验条件下的 CNI 三项参数差异，并计算均值差异的 95%置信区间。结果显示，I 参数存在显著的条件差异：同步条件下被试的一般性不作为偏好(M = 0.622, SD = 0.231)显著低于异步条件(M = 0.713, SD = 0.237)， $t(37) = -2.174$ ， $p = 0.034$ ，Cohen’s $d = 0.367$ ，均值差异(同步 - 异步)为-0.091，95% CI [-0.176, -0.006]。C 参数在两种条件下无显著差异， $t(37) = -0.679$ ， $p = 0.501$ ， $d = 0.110$ ，均值差异的 95% CI [-0.056, 0.028]；N 参数的条件差异同样不显著， $t(37) = -1.024$ ， $p = 0.313$ ， $d = 0.166$ ，95% CI [-0.113, 0.037]。此外，被试对道德决策任务的主观难度评分($p = 0.733$)与情绪卷入评分($p = 0.078$)在两种条件下均无显著差异。具体结果见表 2。

Table 2. Paired-samples t-test results for CNI model parameters under the two conditions (N = 38)

表 2. 两种条件下 CNI 模型参数的配对样本 t 检验结果(N = 38)

参数	组别	平均值	标准差	t	p
C	同步	0.331	0.173	-0.679	0.501
C	异步	0.345	0.154	-0.679	0.501
N	同步	0.368	0.308	-1.024	0.313
N	异步	0.406	0.349	-1.024	0.313
I	同步	0.622	0.231	-2.174	0.034
I	异步	0.713	0.237	-2.174	0.034

4. 讨论

本研究旨在探究机器人化身的身体拥有感错觉对个体道德两难决策的影响。操纵检验结果重复了既往研究结论：个体能够对机器人化身产生全身身体拥有感错觉(Farizon et al., 2021; Alimardani et al., 2013; Aymerich-Franch et al., 2016, 2020)。

需要首先澄清，本研究比较的是同一标准人形机器人化身在同步与异步条件下的差异。因此，条件差异更可能反映身体拥有感错觉或具身同步性，而非机器人相对于人类化身的特异性。若要检验“机器人属性”本身的作用，未来需采用 2(化身类型：机器人/人类)×2(同步/异步)设计，或加入非人客体、无化身对照组。

其次, 本研究发现机器人化身的具身体验显著影响个体的一般性不作为偏好: 相较于成功建立身体拥有感错觉的同步条件, 异步条件(无身体拥有感错觉)下被试表现出更强的不作为偏好, 该结果与研究假设 H2 一致。对此, 至少存在三种可能且彼此并不互斥的解释。其一, 来自社会刻板预期视角: 当个体在同步条件下对机器人化身形成身体拥有感错觉时, 自我表征与机器人躯体发生融合, 个体内化了社会对机器人行动方式的预期——人们往往预期机器人在此类困境中会采取更为结果导向(功利化)的行动(Scheutz & Malle, 2021; Malle et al., 2025)——进而降低自身固有的不作为倾向; 而在异步条件下, 被试以外部观察者的身份看待机器人, 并未发生自我-机器人的表征融合, 因此保留了更强的不作为偏好。

其二, 来自主动控制感(sense of agency)视角: 同步条件通过动作实时跟随赋予被试对机器人躯体的控制感, 表 1 显示主动控制感题目的得分在同步条件下显著高于异步条件; 主动控制感增强使个体更倾向于以行动主体的身份参与情境、主动做出行为, 而非被动旁观或等待, 从而降低不作为偏好。这一解释也与 Farizon 等人(2022)的结论一致——动作同步诱发的主动控制感(而非单纯的视觉具身)才是改变个体对机器人态度与行为倾向的关键。该结果说明, 机器人化身带来的全身具身改变, 能够独立于结果敏感性(C)与规范敏感性(N), 单独调节个体在道德困境中“行动/不行动”的基础行为倾向。

其三, 道德脱离(moral disengagement)也可能是一个候选解释。道德脱离指个体通过责任转移、去人性化等认知机制, 弱化自身行为与道德后果之间的联系, 从而降低实施道德相关行为时的自我制裁与内疚感(Bandura, 1999)。在 AI 道德增强的讨论中, 研究者也指出, 当个体依赖人工智能系统进行道德判断时, 决策责任可能被转移至算法而非自身, 存在诱发道德脱离的风险(Volkman & Gabriels, 2023)。与之类似, 当被试通过身体拥有感错觉真正感知到自己“成为”机器人时, 行动主体的道德身份发生了转移: 决策的执行人被知觉为机器人的躯体, 而不再是具有道德感的“自我”。这种主体与道德自我的分离, 可能使个体产生责任转移与去人性化的认知, 从而出现一定程度的道德脱离。在道德两难情境中, 采取行动通常意味着主动施加伤害、需要承担责任, 会触发强烈的道德厌恶与内疚, 个体因而倾向于不作为; 而一旦发生道德脱离, 主动采取行动所伴随的内疚与责任承担被削弱, 行动造成的道德成本相应降低, 个体便更愿意做出行动选择, 不作为偏好随之下降。该机制与本研究的核心结果相吻合: 仅在成功诱发身体拥有感错觉的同步条件下, 个体“成为”机器人并发生道德脱离, 不作为偏好降低; 而在异步条件下, 被试始终以旁观者视角看待机器人, 未发生自我-机器人的表征融合, 道德脱离程度较低, 因此维持了更高水平的不作为偏好。值得注意的是, 这一解释与 Aymerich-Franch 等人(2020)发现“被试会对机器人化身的失范行为感到内疚并道歉”的结果并不必然矛盾: 其一, 二者考察的行为性质不同, 社交失范行为涉及对他人尊严的冒犯, 而牺牲型道德困境涉及伤害与挽救之间的权衡, 前者更容易触发即时的社会情绪反应; 其二, 内疚的测量多发生在行为事后、被试对化身行为进行评价与回应的阶段, 此时个体的道德自我重新在场, 而道德两难决策要求个体在当下即刻做出取舍, 责任转移更可能在决策发生的当下发挥作用。然而, 必须强调, 本研究未直接测量责任归属、道德脱离、内疚或自我认同, 因此该解释仅为事后假设, 不能作为本研究的结论。未来研究可在决策后立即测量责任归属(如“该决策应由谁负责?”)、道德脱离简版、内疚/羞耻与自我-化身融合, 并检验中介路径。

然而, 实验并未观测到两种条件在结果敏感性(C 参数)与道德规范敏感性(N 参数)上的差异, 该结果与假设 H1 不相符。对此本研究提出两方面解释。第一, 结果敏感性与规范敏感性相比一般性不作为偏好, 属于更为稳定的道德决策倾向, 较难以情境操纵改变(Luke & Gawronski, 2022)。机器人化身带来的具身体验仅改变个体的主体感知, 其干预强度或许不足以撼动被试稳定的结果加工与规范加工模式。需要指出的是, 这一结果与既往 VR 化身道德研究的表现一致: Salagean 等人(2024)发现个性化化身仅提升生理唤醒而不改变道德决策结果, Frisanco 等人(2024)也报告化身上帝对道德两难决策无显著影响。这提示, 化身操纵对道德决策的调节可能主要作用于行动倾向等较为灵活的行为层面, 而难以改变结果权衡

与规范加工等相对稳定的决策模式。第二, 同步条件下的身体拥有感错觉可能通过具身模拟提升个体的共情或情绪卷入。大量研究证实, 具身模拟能够提升个体的共情能力(Seinfeld et al., 2018), 而共情水平越高的个体在面对道德困境时往往拥有更多的情绪反应(Cecchetto et al., 2018)。因此, 同步条件下由具身体验带来的共情增强, 可能使个体对道德困境的情绪反应并未减弱, 甚至有所增强, 从而缓冲了机器人身份可能伴随的情绪抑制效应, 使 N 参数未如 H1 预期下降。本研究中, 同步条件下的情绪卷入评分略高于异步条件, 但差异仅边缘显著($p = 0.078$)。这一方向性趋势与该推测一致, 可视为初步的、边缘性的线索。然而, 情绪卷入评分并非共情的直接测量, 且本研究未测量共情、生理唤醒或内疚等情绪指标, 因此该解释仍属事后假设, 不能作为 N 参数零结果的确定机制。至于结果敏感性(C), 其反映的是个体对行为后果的认知性权衡, 主要依赖于结果信息的可获得性与权衡动机, 而本研究的操纵改变的是行动主体的身体表征而非结果信息本身, 因此 C 参数保持稳定同样处于预期之内。综合来看, 化身具身可能优先改变“是否行动”的倾向性层面, 而对“如何权衡结果与规范”的加工层面影响有限, 这一分离也与 CNI 模型对三类参数相对独立的设定相一致(Gawronski et al., 2017)。

本研究仍存在若干局限。首先, 本研究未设置人类化身对照条件, 因而无法完全排除该效应源于“任何虚拟化具身体验”而非“机器人属性”本身的可能性, 未来可增设人类化身与无化身对照组, 以厘清机器人身份的特异性贡献。其次, 本研究未测量责任归属、道德脱离、内疚、自我认同与共情, 因而无法直接检验责任转移或道德脱离机制。未来研究需采用多化身类型对照, 并加入中介测量。再次, 本研究仅使用一种标准人形机器人, 未系统操纵拟人化程度、面部特征、性别化、机械感与人类相似度。不同形态机器人可能引发不同心理效应, 因此结果不能直接推广到其他机器人化身。最后, 本研究的道德两难任务以文本材料为主, 缺少生理层面的客观指标, 仅依靠行为指标难以充分揭示内部情绪与认知过程, 未来可以结合皮电、心率等生理数据(Salagean et al., 2024; Pucci & Gagliardi, 2025), 对本研究的效应与内在机制进行交叉验证。

随着机器人化身技术逐步融入日常生活, 其对人类道德认知与行为的潜在影响值得学界更多关注。本研究聚焦机器人化身下的道德两难决策, 证实身体拥有感错觉可以降低个体的一般性不作为偏好。该结果也具备一定现实启示: 在未来人机交互的实际场景中, 需要重视化身具身对个体行动倾向的塑造作用, 尽可能对该类心理效应加以考量和干预。

基金项目

2026 年西南大学教育教学改革项目立项(2026JY069)。

参考文献

- 康德(1957). *道德形而上学探本*(唐钺, 译). 北京: 商务印书馆。
- Alimardani, M., Nishio, S., & Ishiguro, H. (2013). Humanlike Robot Hands Controlled by Brain Activity Arouse Illusion of Ownership in Operators. *Scientific Reports*, 3, Article No. 2396. <https://doi.org/10.1038/srep02396>
- Aymerich-Franch, L., Petit, D., Ganesh, G., & Kheddar, A. (2016). The Second Me: Seeing the Real Body during Humanoid Robot Embodiment Produces an Illusion of Bi-Location. *Consciousness and Cognition*, 46, 99-109. <https://doi.org/10.1016/j.concog.2016.09.017>
- Aymerich-Franch, L., Kishore, S., & Slater, M. (2020). When Your Robot Avatar Misbehaves You Are Likely to Apologize: An Exploration of Guilt during Robot Embodiment. *International Journal of Social Robotics*, 12, 217-226. <https://doi.org/10.1007/s12369-019-00556-5>
- Bandura, A. (1999). Moral Disengagement in the Perpetration of Inhumanities. *Personality and Social Psychology Review*, 3, 193-209. https://doi.org/10.1207/s15327957pspr0303_3
- Banakou, D., Kishore, S., & Slater, M. (2018). Virtually Being Einstein Results in an Improvement in Cognitive Task Performance and a Decrease in Age Bias. *Frontiers in Psychology*, 9, Article ID: 917.

- <https://doi.org/10.3389/fpsyg.2018.00917>
- Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting External Validity: Concerns about Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology. *Social and Personality Psychology Compass*, 8, 536-554. <https://doi.org/10.1111/spc3.12131>
- Botvinick, M., & Cohen, J. (1998). Rubber Hands ‘Feel’ Touch That Eyes See. *Nature*, 391, 756-756. <https://doi.org/10.1038/35784>
- Cecchetto, C., Korb, S., Rumiati, R. I., & Aiello, M. (2018). Emotional Reactions in Moral Decision-Making Are Influenced by Empathy and Alexithymia. *Social Neuroscience*, 13, 226-240. <https://doi.org/10.1080/17470919.2017.1288656>
- Dang, J., & Liu, L. (2024). Viewing Machines as Humans but Humans as Machines? Social Connectedness Shapes the Robot Anthropomorphism-Dehumanization Link. *Technological Forecasting and Social Change*, 208, Article 123683. <https://doi.org/10.1016/j.techfore.2024.123683>
- Ehrsson, H. H., Wiech, K., Weiskopf, N., Dolan, R. J., & Passingham, R. E. (2007). Threatening a Rubber Hand That You Feel Is Yours Elicits a Cortical Anxiety Response. *Proceedings of the National Academy of Sciences*, 104, 9828-9833. <https://doi.org/10.1073/pnas.0610011104>
- Farizon, D., Dominey, P. F., & Ventre-Dominey, J. (2021). Insights on Embodiment Induced by Visuo-Tactile Stimulation during Robotic Telepresence. *Scientific Reports*, 11, Article No. 22718. <https://doi.org/10.1038/s41598-021-02091-8>
- Farizon, D., Dominey, P. F., & Ventre-Dominey, J. (2022). Social Attitude Towards a Robot Is Promoted by Motor-Induced Embodiment Independently of Spatial Perspective. *IEEE Robotics and Automation Letters*, 7, 9036-9042. <https://doi.org/10.1109/lra.2022.3189150>
- Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behavior Research Methods*, 39, 175-191. <https://doi.org/10.3758/bf03193146>
- Foglia, L., & Wilson, R. A. (2013). Embodied Cognition. *WIREs Cognitive Science*, 4, 319-325. <https://doi.org/10.1002/wcs.1226>
- Foot, P. (1967). The Problem of Abortion and the Doctrine of the Double Effect. In P. Foot (Ed.), *Virtues and Vices* (pp. 19-32). Oxford University Press. <https://doi.org/10.1093/0199252866.003.0002>
- Frisanco, A., Schepisi, M., Tieri, G., & Aglioti, S. M. (2024). Does Embodying a Divine Avatar Influence Moral Decisions? An Immersive Virtual Reality Study. *Cyberpsychology, Behavior, and Social Networking*, 27, 482-489. <https://doi.org/10.1089/cyber.2023.0242>
- Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, Norms, and Generalized Inaction in Moral Dilemmas: The CNI Model of Moral Decision-Making. *Journal of Personality and Social Psychology*, 113, 343-376. <https://doi.org/10.1037/pspa0000086>
- Gawronski, B., Conway, P., Hütter, M., Luke, D. M., Armstrong, J., & Friesdorf, R. (2020). On the Validity of the CNI Model of Moral Decision-Making: Reply to Baron and Goodwin (2020). *Judgment and Decision Making*, 15, 1054-1072. <https://doi.org/10.1017/s1930297500008251>
- 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, 2105-2108. <https://doi.org/10.1126/science.1062872>
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive Load Selectively Interferes with Utilitarian Moral Judgment. *Cognition*, 107, 1144-1154. <https://doi.org/10.1016/j.cognition.2007.11.004>
- Kilteni, K., Maselli, A., Kording, K. P., & Slater, M. (2015). Over My Fake Body: Body Ownership Illusions for Studying the Multisensory Basis of Own-Body Perception. *Frontiers in Human Neuroscience*, 9, Article ID: 141. <https://doi.org/10.3389/fnhum.2015.00141>
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M. et al. (2007). Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements. *Nature*, 446, 908-911. <https://doi.org/10.1038/nature05631>
- Körner, A., Deutsch, R., & Gawronski, B. (2020). Using the CNI Model to Investigate Individual Differences in Moral Dilemma Judgments. *Personality and Social Psychology Bulletin*, 46, 1392-1407. <https://doi.org/10.1177/0146167220907203>
- Luke, D. M., & Gawronski, B. (2022). Temporal Stability of Moral Dilemma Judgments: A Longitudinal Analysis Using the CNI Model. *Personality and Social Psychology Bulletin*, 48, 1191-1203. <https://doi.org/10.1177/01461672211035024>
- Lu, J., Xie, J., Chen, J., Zeng, Y., Jiang, Z., Wang, Y. et al. (2022). More Utilitarian Judgment in Internet Addiction? An Exploration Using Process Dissociation and the CNI Model. *Brain and Behavior*, 12, e2510. <https://doi.org/10.1002/brb3.2510>
- Malle, B. F., Scheutz, M., Cusimano, C., Voiklis, J., Komatsu, T., Thapa, S. et al. (2025). People’s Judgments of Humans and Robots in a Classic Moral Dilemma. *Cognition*, 254, Article 105958. <https://doi.org/10.1016/j.cognition.2024.105958>
- Mill, J. S. (1992). *On Liberty and Utilitarianism*. Bantam Books.

- Moshagen, M. (2010). Multitree: A Computer Program for the Analysis of Multinomial Processing Tree Models. *Behavior Research Methods*, 42, 42-54. <https://doi.org/10.3758/brm.42.1.42>
- Nishio, S., Watanabe, T., Ogawa, K., & Ishiguro, H. (2012). Body Ownership Transfer to Teleoperated Android Robot. In S. S. Ge, O. Khatib, J. -J. Cabibihan, R. Simmons, & M. -A. Williams (Eds.), *Lecture Notes in Computer Science* (pp. 398-407). Springer. https://doi.org/10.1007/978-3-642-34103-8_40
- Persky, S., & Blascovich, J. (2008). Immersive Virtual Video Game Play and Presence: Influences on Aggressive Feelings and Behavior. *Presence: Teleoperators and Virtual Environments*, 17, 57-72. <https://doi.org/10.1162/pres.17.1.57>
- Petkova, V. I., Khoshnevis, M., & Ehrsson, H. H. (2011). The Perspective Matters! Multisensory Integration in Ego-Centric Reference Frames Determines Full-Body Ownership. *Frontiers in Psychology*, 2, Article No. 35. <https://doi.org/10.3389/fpsyg.2011.00035>
- Pucci, M., & Gagliardi, L. (2025). Moral Dilemmas in Virtual Reality: A Narrative Review of Findings and Methodological Challenges. *Frontiers in Psychology*, 16, Article ID: 1701188. <https://doi.org/10.3389/fpsyg.2025.1701188>
- Salagean, A., Wu, M., Fletcher, G., Cosker, D., & Fraser, D. S. (2024). The Utilitarian Virtual Self—Using Embodied Personalized Avatars to Investigate Moral Decision-Making in Semi-Autonomous Vehicle Dilemmas. *IEEE Transactions on Visualization and Computer Graphics*, 30, 2162-2172. <https://doi.org/10.1109/tvcg.2024.3372121>
- Scheutz, M., & Malle, B. F. (2021). May Machines Take Lives to Save Lives? Human Perceptions of Autonomous Robots (with the Capacity to Kill). In B. J. Strawser (Ed.), *Lethal Autonomous Weapons* (pp. 89-102). Oxford University Press. <https://doi.org/10.1093/oso/9780197546048.003.0007>
- Seinfeld, S., Arroyo-Palacios, J., Iruretagoyena, G., Hortensius, R., Zapata, L. E., Borland, D. et al. (2018). Offenders Become the Victim in Virtual Reality: Impact of Changing Perspective in Domestic Violence. *Scientific Reports*, 8, Article No. 2692. <https://doi.org/10.1038/s41598-018-19987-7>
- Shaheen, T., Brščić, D., & Kanda, T. (2025). Investigation of Low-Moral Actions by Malicious Anonymous Operators of Avatar Robots. *ACM Transactions on Human-Robot Interaction*, 14, 1-34. <https://doi.org/10.1145/3696466>
- Thomson, J. J. (1976). Killing, Letting Die, and the Trolley Problem. *Monist*, 59, 204-217. <https://doi.org/10.5840/monist197659224>
- Volkman, R., & Gabriels, K. (2023). AI Moral Enhancement: Upgrading the Socio-Technical System of Moral Engagement. *Science and Engineering Ethics*, 29, Article 11. <https://doi.org/10.1007/s11948-023-00428-2>
- Yee, N., & Bailenson, J. (2007). The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Communication Research*, 33, 271-290. <https://doi.org/10.1111/j.1468-2958.2007.00299.x>