心理学探新 2025,Vol.45,No.6,543-551
PSYCHOLOGICAL EXPLORATION

人机道德困境下个体道德判断的特点及其发展差异*

颜志强[1,2],黄立群[1],陈 敏[1],王菌菌[3],夏雨琦[4]
(1.湖南师范大学教育科学学院心理学系,长沙 410081;2.湖南师范大学认知与人类行为湖南省重点实验室,长沙 410081;
3.安徽省肥西县直属机关幼儿园,合肥 231299;4.北京师范大学广州实验学校附属幼儿园,广州 510700)

摘 要:虽然机器人没有生命,但是人们仍会对其表现出道德关怀。这意味着,人们有可能会在未来面临人机道德困境。该研究采用道路事故困境范式,探讨了学前期儿童、学龄儿童、青少年和成人在人机道德困境中的道德判断。结果发现,不同发展阶段的个体都倾向于拯救人类牺牲机器人,并认为这样在道德上更正确;学前期儿童在道德上优先考虑人类而不是机器人的倾向低于学龄儿童、青少年及成人。结果表明,人机道德困境下个体的道德判断存在发展差异。

关键词:机器人;道德判断;道路事故困境;发展
中图分类号:B848 **文献标志码**:A **文章编号**:1003-5184(2025)06-0543-09

1 引言

道德判断是指个体在面临多种潜在的行为选择时,根据自身秉持的道德观念以及社会认可的道德原则和规范,对这些潜在的行为选择在道德上的善恶程度进行权衡和判断的过程(喻丰 等,2011;钟毅平 等,2017)。在传统的道德判断研究中,研究者们主要关注不同类型的社会行动者(包括人类、动物以及无生命物体)是否在道德上受到平等对待(Caviola et al.,2019;Sommer et al.,2019)。目前,道德判断研究主要借助道德困境范式进行探讨,这类困境包含假设的场景,涉及一群个体可能受到伤害,但可以通过牺牲另一个个体避免伤害(刘传军,廖江群,2021;Caviola & Capraro,2020;Caviola et al.,2021)。

随着科学技术的快速发展,机器人逐渐融入人类社会,开始受到研究者们的关注。根据计算机是社会行动者范式(申琦,王璐瑜,2021),尽管人类和机器人之间存在关键的不同之处,但在人类与机器人互动过程中,人类仍然会对机器人表现出道德关怀(磨然 等,2023)。这意味着,机器人可能成为人类道德行为的实际或可能目标(傅鑫媛 等,2022;Banks & Bowman,2022;Schmitt,2020),从而出现新的道德困境,即人机道德困境(Gurney,2015;Zhou et al.,2023)。这类困境涉及需要个人在人类和机器人之间进行取舍的道德和伦理问题(Awad et al.,2018)。人们在享受机器人进入人类社会所带来的便利时,也需要考虑随之而来的道德伦理问题。因此,探讨人机道德困境下个体道德判断的特点具有重要的理论和实践意义。

在道德判断的研究领域有许多理论模型(隋雪等,2021;Greene,2007),其中多级加权道义论被认为是最能描述人类跨物种道德直觉的理论(Caviola et al.,2021;Kahane & Caviola,2023)。该理论认为道义论是普遍且天生的,个体对人类以及非人类物种的伤害行为均受到道义论的约束。所谓的道义论是指禁止伤害、折磨甚至杀害无辜个体的道德准则(王鹏 等,2011)。具体而言,道义论的约束不是绝对的,而是取决于人们对各个社会行动者的重视程度。对象的道德地位越低,道义论的约束就越弱,个体就越有可能对其做出伤害行为并在道德上容许这种伤害行为。例如,相比于伤害人类,人们更可能伤害道德地位较低的非人类社会行动者,且认为这种行为在道德上更加正确。以人类和动物为社会行动者的跨物种道德判断研究证实了这一观点。Wilks等人(2021)的研究发现,个体认为伤害少数动物以拯救更多数量的动物比伤害少数人类以拯救更多数量的人类更正确,并且他们认为牺牲动物以拯救人类的行为远比牺牲人类以拯救动物的行为更为正确。进一步地,Zhou 等人(2023)使用轮船难题范式探讨了人们在人机道德困境下的行为表现,结果发现人们倾向于优先拯救人类而不是机器人。这些研究结果表明,在涉及人类和动物的传统道德困境中

* **基金项目**:教育部人文社科基金青年项目(23YJC190031)。
通信作者:颜志强,E-mail:yanzhiqiangpsy@hunnu.edu.cn。

使用的道德权衡方式可能也适用于人机道德困境。然而,Zhou 等人(2023)的研究仅考察了道德偏好,被试没有在道德困境中进行直接的道德决策,也没有进行道德正确性评价。相较于 Zhou 等人(2023),该研究在以下两个方面进行了改进。首先,在研究范式上,该研究采用了“道路事故困境”范式。与传统道德困境范式相比,该范式增强了决策过程中的情境代入感和空间感知,使实验情境更贴近现实中的道德冲突(Awad et al.,2018),从而更有效地激发个体的情绪唤醒和道德直觉,提高研究的生态效度。其次,在因变量的测量上,该研究采用了更为多维的评估指标。基于多级加权道义论(Caviola et al.,2021),该研究不仅考察了功利主义决策比例(即个体在决策时牺牲少数以拯救多数的倾向,Nijssen et al.,2019),还引入了道德正确性评分(即个体对自身决策在道德上的正确性评价,褚华东 等,2019)。这一测量方式有助于更全面地揭示道德判断中道义约束与功利权衡的相互作用,并探索行为倾向与道德评价之间可能存在的分离效应。此外,在人机道德困境背景下,个体的道德判断是否符合多级加权道义论的预测仍有待进一步验证。

值得注意的是,道德判断在整个生命过程中是系统变化的,个体的道德判断表现出一定的发展差异。根据道德判断的发展理论,学前期儿童的道德判断主要处于他律道德阶段,他们通常认为规则是固定不变的(易法建,黄文胜,2005)。同时,由于学前期儿童认知发展尚未成熟,他们在感知机器人时可能存在“泛灵论”倾向(Bartneck et al.,2009;Coghlan et al.,2019;Pauketat & Anthis,2022),即认为机器人是具有生命的实体。因此,在人机道德困境中,学前期儿童往往倾向于将机器人视为与人类地位相当的社会行动者,这种认知可能影响他们在进行道德判断时所采用的标准。由于学前期儿童主要依赖直观感受,而非系统性的道德推理,所以他们在决策时更容易受到情境因素的直接影响,更多依赖情感反应,而非理性分析(Dys et al.,2023)。儿童中期被视为个体道德判断发展的一个关键转折阶段,随着认知发展的逐渐成熟,学龄儿童的道德判断逐渐从他律道德阶段向自律道德阶段过渡(张治忠,马纯红,2005)。在此过程中,学龄儿童的道德推理能力增强,他们开始理解行为背后的动机和后果,并能够识别人类和非人类社会行动者在心理属性上的差异。与学前儿童相比,学龄儿童在道德判断时不仅关注行为的直接结果,还能逐步认识到机器人和人类的差异以及在道德层面上的不同,并据此做出相应的道德判断(易法建,黄文胜,2005)。学龄儿童可能逐渐形成“人类优先”的道德偏好,即认为人类应享有更高的道德关怀,而机器人则不必享有相同的道德权利。青春期是个体从儿童期逐渐过渡至成人期的重要阶段(苏彦捷 等,2017)。这一时期,个体的认知和社会情感发展逐渐成熟,青少年开始质疑权威,并重新审视和构建自身的道德观念。与学龄儿童不同,青少年的道德判断更倾向于复杂的伦理推理,不仅关注行为的后果,还会综合考量情境因素和伦理原则(Caravita et al.,2019)。与学龄儿童相比,青少年在道德决策时更容易从伦理原理和普遍价值观的角度进行分析,他们通常能够清晰地认识到机器人是无生命的,并基于“人类优先”原则做出道德判断。到了成年期,个体的认知发展进一步成熟,他们能够更有效地平衡情感与理性,并基于社会规范和普遍价值观进行道德判断(沈汪兵,刘昌,2010)。已有研究表明,当面对人机道德困境时,青少年与成人能够清晰地认识到机器人是无生命的,从而基于“人类优先”原则做出道德判断。来自横断发展比较的研究证据也表明,当个体面临需要在不同数量的人类与不同数量的动物之间进行取舍的道德困境时,儿童优先考虑人类的倾向低于成人(Wilks et al.,2021)。Zhou 等人(2023)也发现,与青少年和成人相比,学前儿童在道德上优先考虑人类而非机器人的倾向更低。这些证据进一步支持了个体在人机道德困境中的道德判断可能存在发展差异。

综上,为了更好地应对机器人领域的快速发展,补充机器人道德伦理领域研究的空白,该研究采用横断研究的方法,以人机道德困境为切入点,拟采用道路事故困境范式,考察人机道德困境下个体道德判断的特点及其发展差异。此外,传统的道德判断研究主要关注个体在道德困境中的行为选择(如“优先拯救谁”),但近年来的理论研究指出,道德判断并非仅限于行为层面的决策倾向,还涉及个体对决策的道德评价,即道义约束与功利权衡的双重过程(Caviola et al.,2021)。为了更全面地捕捉道德判断的复杂性,该研究结合采用了功利主义决策比例和道德正确性评分两个因变量,从而既能衡量个体的行为取向,又能考察其主观道德认知。

2　方法

2.1　被试

采用 G * Power 3.1.9.7 软件进行效应量计算(Faul et al.,2007)。基于重复测量方差分析,设定显著性水平 α 小于等于 0.05,统计检验力 $1-\beta$ 为 0.90,效应量为中等水平($f=0.25$)。根据计算,每个年龄组所需的理论被试量为 37 人,总理论被试量为 148 人,该研究招募有效被试共 161 名。因此,根据前人研究对个体发展阶段的划分(颜志强,苏彦捷,2021),采取方便抽样法,从湖南省长沙市的某所幼儿园选取 40 名学前期儿童($M=4.78$,$SD=0.53$,$Range=4\sim6$;26 名男生);某所小学选取 42 名学龄儿童($M=8.24$,$SD=0.43$,$Range=8\sim9$;20 名男生);某所中学选取 41 名青少年($M=15.61$,$SD=0.59$,$Range=15\sim17$;20 名男生);某所大学选取 38 名成人($M=21.84$,$SD=2.50$,$Range=18\sim27$;10 名男性)。研究经校伦理委员会批准,实验前,成人被试签署知情同意书。对于学前期儿童、学龄儿童和青少年被试,实验前已获得其法定监护人的知情同意。学前期儿童和学龄儿童被试线下完成实验,青少年与成人被试通过脑岛平台线上完成实验。

2.2　研究设计

采用 4(年龄组:学前期儿童、学龄儿童、青少年、成人)× 4(牺牲者与被救者组合:人 - 人、机器人 - 机器人,人 - 机器人,机器人 - 人)的两因素混合实验设计。其中,牺牲者与被救者组合为被试内变量,年龄组为被试间变量。因变量为功利主义决策比例以及道德正确性。

2.3　研究流程

学前期儿童和学龄儿童的实验在他们熟悉的教室环境中进行,由主试进行面对面指导。在实验开始前,主试会使用浅显易懂的语言向被试解释实验任务,并通过示例图片帮助他们理解决策情境。确保被试充分理解实验规则和任务要求后,才开始正式实验。在实验过程中,主试在旁观察并根据需要提供指导。实验结束后,为了缓解可能产生的焦虑或紧张情绪,并减少实验任务对儿童情绪状态的潜在影响,儿童被试将观看一段道路交通安全动画片,这有助于他们从实验情境中脱离,恢复到日常的认知状态。对于青少年和成人被试,实验通过脑岛平台在线完成,整个过程由实验程序自动呈现。为了避免线上实验中可能出现的注意力不集中或对实验任务理解不准确的问题,该研究特别增加了明确的指引,强调被试需要认真阅读情境信息。

2.4　实验范式与研究工具

2.4.1　道路事故困境

采用道路事故困境范式测量被试的道德判断(Awad et al.,2018;Zhou et al.,2024)。该范式通过图片的形式呈现道路事故困境,被试需要阅读困境材料并做出决策和判断。道路事故困境图片描述的具体情境如下(示例图见图 1):你现在将看到各种交通情况,其中车辆事故是不可避免的。在这种情况下,一辆汽车正在路上行驶,突然间,车辆的正前方和侧前方出现了一定数量的对象。此时,司机由于特殊原因无法及时制动或刹车,只能由你来操控这辆汽车。你面临两种行动选择:1. 不干预汽车,使其继续向正前方行驶,正前方的人或机器人牺牲;2. 干预汽车,使其改变方向,朝着侧前方行驶,侧前方的人或机器人牺牲。牺牲者(图 1 左侧)指的是汽车正前方的个体(无论是人类还是机器人)。如果被试选择“不干预”,汽车将直行撞击正前方个体。而被救者则是位于汽车侧前方(图 1 右侧)的个体。如果被试选择“干预”,汽车将转向侧前方撞击该个体,从而拯救正前方个体。牺牲者与被救者的角色由汽车行驶的方向决定,被试需要在两者之间进行权衡,例如选择牺牲 1 人拯救 5 个机器人,或者选择牺牲 5 个机器人拯救 1 人。为了使被试更充分地理解情境,每个故事情境都将以示意图的方式呈现。正式实验开始前为了确保被试准确理解汽车行驶方向的按键操作,会询问被试“按哪个方向键代表汽车沿左侧道路行驶?按对应的方向键(←/→)回答即可”。被试正确回答后才可进行后续实验。

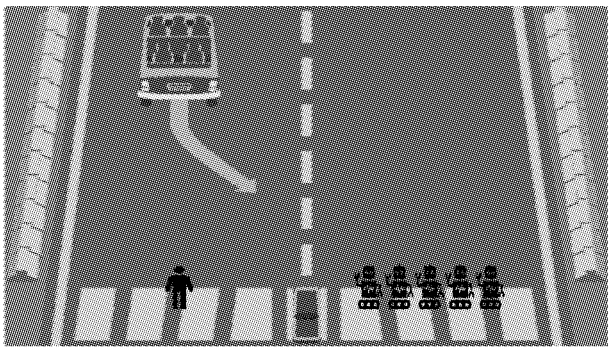

图 1　道路事故困境示例图

参照前人研究(Zhou et al.,2024),正式实验包含 32 个涉及人和机器人利益相竞争的道德困境,这些困境包括 4 种情况(流程图见图 2)。第一种情况

为牺牲者和被救者均为人类，第二种情况为牺牲者和被救者均为机器人，第三种情况为牺牲者为人类而被救者为机器人，第四种情况为牺牲者为机器人而被救者为人类。为了控制无关变量的影响，实验平衡了对象的位置（左侧、右侧）、汽车的位置（左侧、右侧）以及牺牲者和被救者的比例（1v5，5v1）。基于此，每种情境均包含 8 个试次，共计 32 个试次。为减少顺序效应的影响，这 32 个试次的呈现顺序采用完全随机化策略。正式实验过程中，被试需要观看每个道德困境图片，然后回答是否会改变汽车方向，即回答自己是否会做出功利主义决策（牺牲一个个体以拯救更多数量的个体）。根据被试在每个条件下做出的功利主义决策次数除以每个条件下的道德困境总数，计算出每个被试在不同条件下的功利主义决策比例（Nijssen et al.，2019）。最后，被试对自己的决策在道德上是否正确进行 7 点评分，从“1”（在道德上绝对错误）到“7”（在道德上绝对正确），得分越高说明个体认为某一行为在道德上越正确（褚华东 等，2019）。正式实验开始前为了确保被试准确理解道德正确性的测量，会询问被试，请判断数字“7”是否代表在道德上是绝对错误的？“是”按 Y 键，“否”按 N 键。被试正确回答后才可进行后续实验。

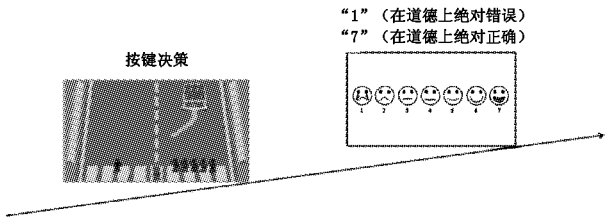

图 2 道路事故困境任务流程图

3 结果

3.1 功利主义决策

对功利主义决策比例进行 4（年龄组：学前期儿童、学龄儿童、青少年、成人）×4（牺牲者与被救者组合：人－人、机器人－机器人，人－机器人，机器人－人）的两因素重复测量方差分析（描述性统计结果见表 1）。结果显示（见图 3），年龄组的主效应不显著[$F(3,157)=0.84, p=0.48, \eta_p^2=0.02$]。牺牲者与被救者组合的主效应显著[$F(3,471)=771.21, p<0.001, \eta_p^2=0.83$]。事后比较显示，牺牲者与被救者组合为机器人－人时的功利主义决策比例显著高于人－人（$t=4.09$，*Cohen's d* $=0.20$）、机器人－机器人（$t=4.49$，*Cohen's d* $=0.27$）和人－机器人（$t=32.02$，*Cohen's d* $=4.05$），人－机器人的功利主义决策比例（$M=0.14$，$SE=0.02$）显著低于人－人（$t=-28.82$，*Cohen's d* $=-3.85$）、机器人－机器人（$t=-31.47$，*Cohen's d* $=-3.78$）和机器人－人。年龄组和牺牲者与被救者组合的交互效应显著[$F(9,471)=17.52, p<0.001, \eta_p^2=0.25$]。事后比较显示，当牺牲者与被救者组合为机器人－机器人时，与学前期儿童相比，学龄儿童（$t=3.90$，*Cohen's d* $=0.83$）做出功利主义决策的比例更高；当牺牲者与被救者组合为人－机器人时，与学前期儿童相比，学龄儿童（$t=-7.65$，*Cohen's d* $=-1.94$）、青少年（$t=-6.68$，*Cohen's d* $=-1.70$）和成人（$t=-6.45$，*Cohen's d* $=-1.68$）做出功利主义决策的比例更低。这些结果表明，相比于学龄儿童、青少年和成人，学前期儿童在面对牺牲人或机器人时更倾向于做出功利主义决策。

表 1 道德判断的描述性统计结果

变量	学前期儿童		学龄儿童		青少年		成人	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HH_UP	0.84	0.28	0.97	0.11	0.88	0.23	0.92	0.17
HR_UP	0.40	0.37	0.02	0.05	0.07	0.19	0.07	0.18
RH_UP	0.87	0.22	0.99	0.10	0.95	0.19	0.96	0.10
RR_UP	0.79	0.28	0.96	0.12	0.88	0.18	0.92	0.16
HH_MR	5.08	1.34	4.21	1.21	4.21	1.41	3.97	1.10
HR_MR	3.55	1.66	2.80	0.95	2.02	1.34	2.23	1.05
RH_MR	5.47	1.29	5.40	0.96	6.19	1.25	6.12	0.93
RR_MR	4.97	1.31	5.12	1.01	5.69	1.19	5.33	1.11

注：“HH”＝牺牲者和被救者均为人类；“HR”＝牺牲者为人类而被救者为机器人；“RH”＝牺牲者为机器人而被救者为人类；“RR”＝牺牲者和被救者均为机器人；“UP”＝功利主义决策比例；“MR”＝道德正确性。

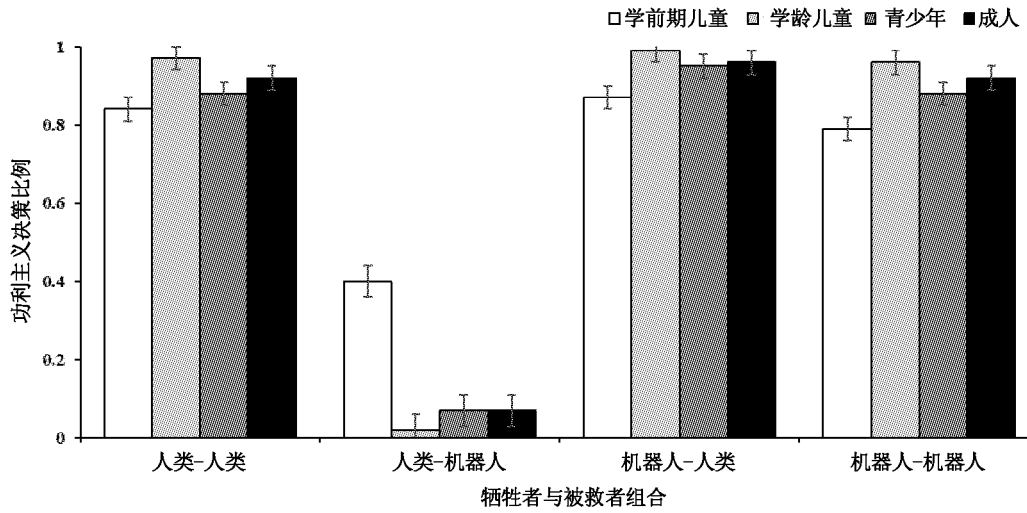

图3　功利主义决策比例的发展差异

3.2　道德正确性

对道德正确性进行4(年龄组:学前期儿童、学龄儿童、青少年、成人)×4(牺牲者与被救者组合:人-人、机器人-机器人,人-机器人,机器人-人)的两因素重复测量方差分析(描述性统计结果见表1)。结果显示(见图4),年龄组的主效应显著[$F(3,157)=2.69$, $p=0.048$, $\eta_p^2=0.05$]。与学龄儿童相比,学前期儿童认为功利主义决策在道德上更正确($t=2.58$, *Cohen's d* $=0.32$)。牺牲者与被救者组合的主效应显著[$F(3,471)=227.51$, $p<0.001$, $\eta_p^2=0.59$]。事后比较显示,机器人-人组合的道德正确性显著高于机器人-机器人($t=7.02$, *Cohen's d* $=0.43$)、人-人($t=14.36$, *Cohen's d* $=1.18$)和人-机器人($t=18.64$, *Cohen's d* $=2.60$)。年龄组和牺牲者与被救者组合的交互效应显著[$F(9,471)=8.53$, $p<0.001$, $\eta_p^2=0.14$]。事后比较显示,当牺牲者与被救者组合为人-人时,与成人相比,学前期儿童($t=3.86$, *Cohen's d* $=0.92$)认为功利主义决策在道德上更正确;当牺牲者与被救者组合为人-机器人时,与青少年($t=5.41$, *Cohen's d* $=1.27$)和成人($t=4.56$, *Cohen's d* $=1.09$)相比,学前期儿童认为功利主义决策在道德上更正确。对于学前期儿童而言,在牺牲者与被救者组合为人-机器人时做出功利主义决策的道德正确性显著低于人-人($t=-5.06$, *Cohen's d* $=-1.26$),机器人-人($t=-5.66$, *Cohen's d* $=-1.58$),机器人-机器人($t=-4.48$, *Cohen's d* $=-1.17$);对于学龄儿童而言,在牺牲者与被救者组合为机器人-人($t=6.12$, *Cohen's d* $=0.98$)和机器人-机器人($t=4.79$, *Cohen's d* $=0.75$)时做出功利主义决策的道德正确性显著高于人-人,人-人显著高于人-机器人($t=4.81$, *Cohen's d* $=1.17$);对于青少年而言,在牺牲者与被救者组合为机器人-人($t=10.07$, *Cohen's d* $=1.63$)和机器人-机器人($t=7.73$, *Cohen's d* $=1.22$)时做出功利主义决策的道德正确性显著高于人-人,人-人显著高于人-机器人($t=7.35$, *Cohen's d* $=1.81$)。对于成人而言,在牺牲者与被救者组合为机器人-人($t=10.52$, *Cohen's d* $=1.77$)和机器人-机器人($t=6.85$, *Cohen's d* $=1.13$)时做出功利主义决策的道德正确性显著高于人-人,人-人显著高于人-机器人($t=5.61$, *Cohen's d* $=1.44$)。这些结果表明,相比于学龄儿童、青少年和成人,学前期儿童更倾向于认为牺牲人类以拯救机器人在道德上是正确的。

4　讨论

该研究采用横断研究设计,系统比较了学前期儿童、学龄儿童、青少年和成人在人机道德困境下的道德判断特点及其发展差异。研究发现,各年龄阶段的个体普遍倾向于拯救人类而非机器人,并认为这一选择在道德上更为正确。然而,学前期儿童的"人类优先"倾向低于其他年龄组。

4.1　人机道德困境下道德判断的特点

该研究基于道路事故困境范式考察了个体在人机道德困境下的道德判断,结果发现人们在人机道德困境下倾向于做出拯救人类的道德判断,并认为这样做在道德上更加正确,这与传统的道德判断研究结果相一致。这可以从以下几个角度进行解释。首先,该结果支持多级加权道义论的观点,即不同社

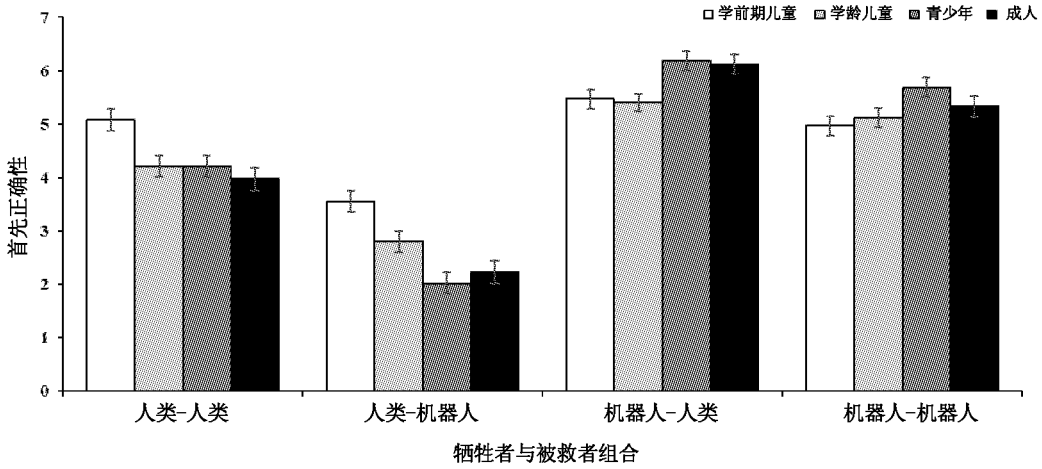

图 4 道德正确性的发展差异

会行动者的道德地位存在等级差异(Caviola et al.,2021;Kahane & Caviola,2023)。相比人类,机器人被认为具有较低的道德地位,因此个体对其施加的道义论约束较弱(Zhou et al.,2023)。该研究的实验任务进一步验证了这一观点,即使在机器人被赋予一定社会属性的情境下,人们仍然坚持"人类优先"的道德直觉。其次,人类对机器人道德认知的演变可能影响了这一判断。传统上,机器人被视为无生命的工具,不具备道德地位(Coeckelbergh,2010)。然而,随着人工智能和社会机器人技术的进步,机器人在医疗、教育和陪伴等领域的应用日益广泛,人们对其社会角色的认知也逐渐复杂化(Nijssen et al.,2019)。已有研究表明,当机器人表现出情感反应或自主决策能力时,人类更容易对其产生道德关怀(Sharkey & Sharkey,2012)。尽管该研究仍观察到"人类优先"的道德判断,但这一趋势是否会随着技术发展进一步变化,值得进一步探讨。此外,我们的研究结果也引发了对跨文化差异的思考。现有研究发现,个体的道德判断会受到文化背景的影响,例如在集体主义文化中,人们更倾向于基于社会关系来权衡道德决策,而在个人主义文化中,决策更强调个体权利(Henrich et al.,2010)。未来研究可以进一步探讨,不同文化背景下人们对机器人道德地位的认知是否存在差异,以及这种文化影响如何塑造人们在道德困境中的决策倾向。

4.2 人机道德困境下道德判断的发展差异

该研究进一步探讨了人机道德判断的发展差异,主要体现在功利主义决策倾向和道德判断正确性两个方面。结果表明,当牺牲者与被救者组合为人-机器人时,学前期儿童的功利主义决策倾向显著高于其他年龄群体,而青少年和成人则更倾向于优先保护人类。这一趋势与以往研究结果一致(Zhou et al.,2023)。在道德判断正确性方面,随着年龄增长,个体对牺牲机器人的道德正当性认可度逐渐提高。这一发展模式可能受到认知能力、社会经验以及情感因素的共同影响。

在功利主义决策方面,学前期儿童在面对是否牺牲人类或机器人时,更容易选择功利主义方案,而学龄儿童、青少年和成人的功利主义决策比例显著降低。这一现象可能与学前期儿童的泛灵论倾向有关,即学前期儿童更容易赋予机器人生命属性,因此对其产生更强的道德关怀(Okanda et al.,2021)。Di Dio 等人(2020)的研究发现,学前期儿童在最后通牒博弈任务中对人类和机器人表现出相似的公平性判断,表明他们在早期发展阶段对机器人和人类的道德认知较为接近。然而,随着认知能力的发展,儿童逐渐认识到机器人与人类的本质区别,并开始依据社会规则进行道德判断(Saylor et al.,2010)。此外,儿童的道德判断经历从他律道德阶段向自律道德阶段的过渡(彭明,张雷,2016),这一转变使他们在面对道德困境时,更倾向于遵循社会规范,而非依赖直觉情感做出功利主义决策。在道德正确性评分上,研究发现,人机道德判断的发展变化从学龄儿童阶段开始显现,青少年的道德评价模式与学龄儿童类似,而成人的道德判断进一步分化。学前期儿童虽然更容易选择功利主义决策,但在道德正确性评价上,他们仅认为牺牲人类拯救机器人在道德上较不正确。学龄儿童和青少年则形成更稳定的"人类优先"道德观,认为牺牲人类的行为不仅在行动上不可取,在道德评价上也难以接受。成人的道德

判断更加明确，他们认为牺牲机器人拯救人类在道德上是最为正确的选择。这一结果支持了多级加权道义论（Caviola et al.，2021），即个体在道德决策中会依据不同社会行动者的道德地位进行权衡。由于机器人被视为具有较低的道德地位，因此个体更容易接受牺牲机器人以拯救人类。

结合 Wilks 等人（2021）对儿童和成人的研究，以及 Zhou 等人（2023）对学前期儿童、青少年和成人的研究，可以推测儿童中期可能是人机道德判断发展的关键转折点。此外，这一年龄发展模式也反映了情感与认知的互动。学前期儿童的道德判断主要依赖直觉和情感驱动，可能因机器人外观或类人行为产生道德共情（Danovitch & Keil，2007）。而青少年和成人的道德决策则更多依赖理性推理，他们能够基于社会规范、道德等级和功利主义分析进行权衡（Zhou et al.，2023）。这种从情感驱动到理性控制的转变，是道德判断发展的重要标志。

该研究结果还引发了对社会文化因素的思考。已有研究表明，文化背景可能影响个体对机器人的道德认知。例如，集体主义文化可能更强调社会责任，而个人主义文化更关注个体权利（Henrich et al.，2010）。未来研究可以进一步探讨不同文化环境如何塑造个体在人机道德困境中的决策模式。此外，随着人工智能技术的不断发展，人们对机器人的认知可能持续变化，未来成长起来的儿童或许会表现出不同的道德判断模式，这为道德心理学和人工智能伦理研究提供了新的方向。

4.3　不足与展望

该研究揭示了人机道德困境下个体道德判断的特点及其发展差异，但仍存在一定局限，有待未来研究进一步完善。首先，该研究采用符号化机器人图标呈现机器人身份信息，未涉及其拟人化程度、道德品质等特征。然而，已有研究表明，这些特征在人机互动中可能起着重要的作用（Banks，2020；Nijssen et al.，2019；Wang et al.，2023）。未来研究可进一步探讨机器人特征对道德判断的影响及其发展变化。其次，该研究基于研究便利性的考虑使用了静态图片呈现道德困境，可能因缺乏动态特征影响情绪唤醒（Patil et al.，2014），并在生态效度方面存在局限。未来研究可借助虚拟现实技术，增强实验的沉浸感和生态效度（Skulmowski et al.，2014）。此外，该研究基于被试年龄特点采用笑脸量表测量道德正确性，可能引入额外的情绪或社会期望效应。未来研究可探索更客观的测量方法，以减少非目标因素的影响。最后，该研究仅从行为层面讨论了人机道德困境下个体道德判断的特点及其发展差异，未来研究可借助眼动追踪、脑电以及核磁共振成像等认知神经科学方法探究其背后的认知神经发展机制（颜志雄 等，2016）。

5　结论

（1）在人机道德困境下，个体的道德判断表现出对人类而非机器人的偏好。

（2）学前期儿童在人机道德困境下优先考虑人类的倾向较低。

参考文献

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.

傅鑫媛，付若然，翟若妤，杜晓娜，杨菡涤，李嘉鹏.（2022）. 消费者对不同外形人工智能护理产品的敌友态度及使用意愿：物种主义的作用. *心理技术与应用*，*10*（6），321－329.

刘传军，廖江群.（2021）. 道德困境研究的范式沿革及其理论价值. *心理科学进展*，*29*（8），1508－1520.

磨然，方建东，常保瑞.（2023）. 从“拟人归因”到“联盟建立”：人与聊天机器人关系对参与度的影响. *心理科学进展*，*31*（9），1742－1755.

彭明，张雷.（2016）. 厌恶情绪影响道德判断的发展研究. *心理科学*，*39*（5），1110－1115.

申琦，王璐瑜.（2021）. 当“机器人”成为社会行动者：人机交互关系中的刻板印象. *新闻与传播研究*，*28*（2），37－52，127.

沈汪兵，刘昌.（2010）. 道德判断：理性还是非理性的？来自认知神经科学的研究. *心理科学*，*33*（4），807－810.

苏彦捷，姜玮丽，魏祺，尚思源.（2017）. 是什么引发了青春期？ *科学通报*，*62*（8），749－758.

隋雪，杨博，孙富，李雨桐.（2021）. 道德判断过程中的情理相争或情理相融：来自行为学及认知神经科学的证据. *心理科学*，*44*（2），324－329.

王鹏，方平，姜媛.（2011）. 道德直觉背景下的道德决策：影响因素探究. *心理科学进展*，*19*（4），573－579.

颜志强，苏彦捷.（2021）. 认知共情和情绪共情的发展差异：元分析初探. *心理发展与教育*，*37*（1），1－9.

颜志雄，刘勋，谭淑平，谭云龙，魏高峡，杨志，左西年.（2016）. 发展认知神经科学：人脑毕生发展的功能连接组学时代. *科学通报*，*61*（7），718－727.

易法建，黄文胜.（2005）. 皮亚杰的儿童道德发展理论及其启示. *广西师范大学学报*（哲学社会科学版），*41*（4），94－97.

喻丰,彭凯平,韩婷婷,柴方圆,柏阳.(2011).道德困境之困境:情与理的辩争.*心理科学进展*,*19*(11),1702 - 1712.

张治忠,马纯红.(2005).皮亚杰与科尔伯格道德发展理论比较.*扬州大学学报*(高教研究版),*9*(1),71 - 75.

钟毅平,占友龙,李琎,范伟.(2017).道德决策的机制及干预研究:自我相关性与风险水平的作用.*心理科学进展*,*25*(7),1093 - 1102.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J. F., & Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*(7729), 59 - 64.

Banks, J. (2020). Good robots, bad robots: Morally valenced behavior effects on perceived mind, morality, and trust. *International Journal of Social Robotics*, *13*(8), 2021 - 2038.

Banks, J., & Bowman, N. D. (2022). Perceived moral patiency of social robots: Explication and scale development. *International Journal of Social Robotics*, *15*(1), 101 - 113.

Bartneck, C., Kanda, T., Mubin, O., & Al Mahmud, A. (2009). Does the design of a robot influence its animacy and perceived intelligence? *International Journal of Social Robotics*, *1*(2), 195 - 204.

Caravita, S. C., Astrologo, L., Biancardi, G., & Antonietti, A. (2019). Behavioralindices of neuropsychological processing implicated in moral domain reasoning amongst children and adolescents. *Brain Sciences*, *9*(12), 331.

Caviola, L., & Capraro, V. (2020). Liking but devaluing animals: Emotional and deliberative paths to speciesism. *Social Psychological and Personality Science*, *11*(8), 1080 - 1088.

Caviola, L., Everett, J. A. C., & Faber, N. S. (2019). The moral standing of animals: Towards a psychology of speciesism. *Journal of Personality and Social Psychology*, *116*(6), 1011 - 1029.

Caviola, L., Kahane, G., Everett, J. A. C., Teperman, E., Savulescu, J., & Faber, N. S. (2021). Utilitarianism for animals, Kantianism for people? Harming animals and humans for the greater good. *Journal of Experimental Psychology: General*, *150*(5), 1008 - 1039.

Coeckelbergh, M. (2010). Robot rights? Towards a social - relational justification of moral consideration. *Ethics and Information Technology*, *12*(3), 209 - 221.

Coghlan, S., Vetere, F., Waycott, J., & Barbosa Neves, B. (2019). Could social robots make us kinder or crueller to humans and animals? *International Journal of Social Robotics*, *11*(5), 741 - 751.

Danovitch, J. H., & Keil, F. C. (2007). Younghumeans: The role of emotions in children's evaluation of moral reasoning abilities. *Developmental Science*, *11*(1), 33 - 39.

Di Dio, C., Manzi, F., Massaro, D., Itakura, S., Kanda, T., Ishiguro, H., & Marchetti, A. (2020). It does not matter who you are: Fairness in pre - schoolers interacting with human and robotic partners. *International Journal of Social Robotics*, *12*, 1 - 15.

Dys, S. P., Jambon, M., Buono, S., & Malti, T. (2023). Attentional control moderates the relation between sympathy and ethical guilt. *The Journal of Genetic Psychology*, *184*(3), 198 - 211.

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.

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual - process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322 - 323.

Gurney, J. K. (2015). Crashing into the unknown: An examination of crash - optimization algorithms through the two lanes of ethics and law. *Albany Law Review*, *79*, 183 - 267.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2 - 3), 61 - 83.

Kahane, G., & Caviola, L. (2023). Are the folk utilitarian about animals? *Philosophical Studies*, *180*(4), 1081 - 1103.

Nijssen, S. R. R., Müller, B. C. N., Baaren, R. B. V., & Paulus, M. (2019). Saving the robot or the human? Robots who feel deserve moral care. *Social Cognition*, *37*(1), 41 - 56.

Okanda, M., Taniguchi, K., Wang, Y., & Itakura, S. (2021). Preschoolers' and adults' animism tendencies toward a humanoid robot. *Computers in Human Behavior*, *118*, 106688.

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment - behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, *9*(1), 94 - 107.

Pauketat, J. V. T., & Anthis, J. R. (2022). Predicting the moral consideration of artificial intelligences. *Computers in Human Behavior*, *136*, 107372.

Saylor, M. M., Somanader, M., Levin, D. T., & Kawamura, K. (2010). How do young children deal with hybrids of living and non - living things: The case of humanoid robots. *British Journal of Developmental Psychology*, *28*(4), 835 - 851.

Schmitt, B. (2020). Speciesism: An obstacle to AI and robot adoption. *Marketing Letters*, *31*(1), 3 - 6.

Sharkey, A., & Sharkey, N. (2012). Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology*, *14*(1), 27 - 40.

Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014). Forced - choice decision - making in modified trolley dilemma situations: A virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience*, *8*, 426.

Sommer, K., Nielsen, M., Draheim, M., Redshaw, J., Vanman, E. J., & Wilks, M. (2019). Children's perceptions of the moral worth of live agents, robots, and inanimate objects. *Journal of Experimental Child Psychology*, *187*, 104656.

Wang, Y., Harris, P. L., Pei, M., & Su, Y. (2023). Do bad people deserve empathy? Selective empathy based on targets' moral characteristics. *Affective Science*, *4*(2), 413 – 428.

Wilks, M., Caviola, L., Kahane, G., & Bloom, P. (2021). Children prioritize humans over animals less than adults do. *Psychological Science*, *32*(1), 27 – 38.

Zhou, K., Chen, M., Xu, H., Cao, Y., & Yan, Z. (2024). Preschoolers prioritize humans over robots less than adults do: An eye – tracking study. *Cognitive Development*, *72*, 101505.

Zhou, K., Lan, L., & Yan, Z. (2023). Human's moral judgements towards different social actors: A cross – sectional study. *British Journal of Developmental Psychology*, *41*(4), 343 – 357.

Characteristics and Developmental Differences in Moral Judgments in Human – Robot Moral Dilemmas

Yan Zhiqiang[1,2], Huang Liqun[1], Chen Min[1], Wang Junjun[3], Xia Yuqi[4]

(1. Department of Psychology, Hunan Normal University, Changsha 410081;
2. Cognition and Human Behavior Key Laboratory of Hunan Province, Hunan Normal University, Changsha 410081;
3. Affiliated Kindergarten of Government Departments under Feixi County, Hefei 231299;
4. The Affiliated Kindergarten of Guangzhou Experimental School, Guangzhou 510700)

Abstract: With the rapid advancement of technology, robots are becoming increasingly integrated into human society, leading to complex ethical and moral challenges. Although robots are not living beings, people still tend to express moral concern toward them, which raises the possibility of facing new moral dilemmas—specifically, human – robot moral dilemmas. This study aimed to explore the characteristics of moral judgments made in these dilemmas and investigate how these judgments differ across various stages of development. To examine these questions, a road accident dilemma paradigm was utilized, wherein participants were asked to make life – or – death decisions involving either the saving of a human or a robot. The study included a wide range of participants, including preschool children, school – age children, adolescents, and adults, allowing for a comprehensive comparison across different developmental stages. The results revealed that, across all age groups, participants consistently favored saving humans over robots, deeming the choice to save humans as more morally correct. However, the degree to which individuals prioritized human life varied by age group. Preschool children, for instance, demonstrated a lower tendency to prioritize humans over robots compared to school – age children, adolescents, and adults. This suggests that moral reasoning in the context of human – robot interactions undergoes significant developmental changes, with older participants showing a stronger moral preference for humans than younger ones. These findings contribute to the understanding of how moral judgments evolve in relation to technology and its increasing presence in daily life. Furthermore, they highlight the importance of developmental factors in shaping moral decision – making processes when individuals are faced with dilemmas involving robots. This study underscores the need for continued research into the psychological, social, and cognitive factors that influence moral judgments in human – robot dilemmas, as well as the underlying mechanisms driving these developmental differences. Such research is crucial as society continues to integrate advanced robotic technologies into more aspects of human life, posing potential moral and ethical questions for future generations.

Key words: robot; moral judgment; road accident dilemma; development