

# 感知机器人威胁降低亲社会倾向\*

许丽颖 张语嫣 喻丰

(武汉大学心理学系, 武汉 430072)

**摘要** 机器人进入社会可能会对人类造成心理威胁,而这种威胁会给人际关系带来挑战。通过8个研究,结合档案数据库回溯、问卷调查、情境实验和线下调查,文章探讨了感知机器人威胁对亲社会倾向的影响及其内在机制和边界条件。研究发现:感知机器人威胁会降低人们的亲社会倾向(研究1~7);其机制为集体焦虑的中介作用,即感知机器人威胁会增加集体焦虑,从而降低亲社会倾向(研究2~4);此效应受到内外群体的调节,即感知机器人威胁主要会降低对于外群体成员的亲社会倾向(研究5);同时,此效应受到道德比较倾向的调节,即感知机器人威胁主要会降低下行道德比较者的亲社会倾向(研究6)。研究结果揭示了感知机器人威胁对人际关系的负面影响,拓展了现有关于机器人社会影响的研究。

**关键词** 亲社会倾向, 感知机器人威胁, 集体焦虑, 外群体, 道德比较

**分类号** B849:C91

## 1 引言

人们对技术失业(technological unemployment)的担心自工业革命开始便从未消失,从过去的机器到现在的人工智能机器人,人们一直担心其通过习得人类技能而取代人类劳动(Susskind & Susskind, 2015)。现今人工智能技术发展对于特定模式的常规工作任务冲击更大(Neufeind et al., 2018),若仅看数据,人工智能在就业市场上已经对人类产生严重威胁,甚至有预测表明,美国47%的现有工作将被人工智能取代,这一比例在欧洲的研究中更高达54%(Frey & Osborne, 2017)。人工智能带来的劳动力变革会使简单无认知渗入的重复劳动成为机器人的最佳适用场景,从而导致大部分工人成为“无用阶层”(Harari, 2017)。关于人工智能对不同职业取代程度的讨论不绝于耳,甚至有学者计算出各行业被人工智能替代的不同比例供求职者参考(Frey & Osborne, 2017)。然而,机器人可能对人类造成的威胁还不止于此。除了抢占人类工作岗位等所造成的现实威胁(realistic threat),在外观、能力等方面日

趋拟人化的机器人还可能会对人类造成认同威胁(identity threat),即模糊人与机器之间的界限,从而威胁人类身份认同以及人之为人的独特性(Yogeeswaran et al., 2016)。

尽管机器人技术的发展正在深刻改变全球经济(Agrawal et al., 2019; Ford, 2009),但人们对机器人威胁的感知可能会产生重要的社会影响,尤其是在影响人际关系方面,类似于早期工业发展所引发的社会混乱。现实群际冲突理论认为,日益自动化将加剧群体间的紧张关系,因为它增加了群体之间对资源争夺的竞争(Bobo, 1983; Jackson, 1993)。例如,工业革命放大了欧洲和美国基于种族和阶级的敌对情绪(Mathias, 2013)。在这些历史与理论背景下,最近的研究开始关注机器人威胁对人类社会的影响。例如,与机器人的接触增加对反移民政策的支持(Frey et al., 2018);对自动化威胁的感知会增加人们对限制性移民政策的支持,导致对移民更多的歧视行为(Gamez-Djokic & Waytz, 2020)。近期也有研究者发现了机器人在工作场所的负面人际影响,如,工作场所机器人的增加会造成人们更强的

收稿日期: 2023-12-02

\* 国家社科基金青年项目(20CZX059)。

通信作者: 喻丰, E-mail: psychpedia@whu.edu.cn

工作不安全感,这与工作倦怠及职场不文明行为呈显著正相关(Yam et al., 2023);人工智能的部署增加了工作场所员工的不道德行为(Bai et al., 2025);感知机器人威胁增加职场物化(许丽颖等, 2024)等。这些都反映了机器人技术不仅会影响人际社会,还可能带来负面后果。

在这样的背景下,研究机器人威胁对人们亲社会倾向的影响变得尤为重要。亲社会行为是指个体主动做出使他人、团体或社会受益的行为,旨在促进他人而非自身利益(Batson & Powell, 2003)。亲社会行为因其对人际关系和社会整体福祉的重要意义受到广泛关注(Penner et al., 2005)。现有文献探讨了个体内部因素(如人格、情绪、动机和偏好)与外部因素(如规范、声誉和关系)对亲社会行为的影响(Simpson & Willer, 2015)。然而,随着人工智能技术的迅猛发展和普及,机器人逐渐出现在各种环境中,如工作、家庭和服务场所等。其对人们产生的威胁感知可能成为影响亲社会行为的新的外部因素,并导致人际关系恶化和社会冷漠。基于此,本研究意在探讨感知机器人威胁对亲社会倾向的影响,并试图揭示其背后的潜在机制和边界条件。这将有助于在机器人技术日益普及的背景下,更好地理解 and 应对机器人带来的对人际关系的挑战。

### 1.1 感知机器人威胁如何影响亲社会倾向

机器人技术的发展给社会带来了深刻的变革(Agrawal et al., 2019; Ford, 2009),但同时也引发了人们对机器人威胁的深切焦虑。关于机器人威胁,人们常常描绘着一幅黯淡的未来图景。例如,人们担心机器人会取代自己的工作,或导致人类的未来变得不确定和充满危机(Asimov, 1950; Gray et al., 2023)。这些担忧不仅存在于个体层面,也在集体层面引发广泛焦虑。

个体层面的焦虑通常指对个人生活和未来的不确定性(Mann et al., 2020; Miloyan et al., 2018),例如对职业生涯和经济安全的担忧(Li et al., 2019)。集体层面的焦虑则涉及对整个群体未来的广泛危机感(Wohl & Branscombe, 2008)。关于机器人威胁导致个体焦虑及其对人际关系的影响已经得到了广泛研究(Frey et al., 2017; Im et al., 2019; Wu et al., 2024; Yam et al., 2023)。例如,在职场中接触机器人会引发工作不安全感,并导致对同事的不文明行为(Yam et al., 2023)。自动化对未来就业的威胁会增加人们的物质不安全感,并增加他们对反移民政策的支持(Frey et al., 2017; Im et al., 2019)。

目前,关于机器人威胁引发的个体相关焦虑及其负面人际后果的研究已经较为清晰(Frey et al., 2017; Im et al., 2019; Wu et al., 2024; Yam et al., 2023)。然而,与其他威胁来源类似(Kahn et al., 2022),机器人威胁也可能引发集体焦虑。集体焦虑(collective angst)是一种对群体未来命运的担忧,反映出群体成员对未来的不安情绪(Wohl & Branscombe, 2008)。

集体焦虑有两种来源,即群体的独特性威胁和灭绝威胁(Wohl et al., 2012),这类似机器人给人带来的认同威胁与现实威胁(Yogeeswaran et al., 2016)。首先,根据社会认同理论(social identity theory),群体成员有保持自身群体独特性的动机(Brewer, 2001; Tajfel & Turner, 1986),当人们认为自己所在群体在未来可能会失去与另一个群体相比较时的独特性时,就会产生集体焦虑。这表现在人们担心庆祝他族节日及采用他族仪式和术语会丧失传统、少数民族认为与他族通婚会导致文化泯灭,加拿大人担心在全球化浪潮中变得过于英美化等(Lalonde, 2002; Roccas & Amit, 2011; Wohl et al., 2011)。这种独特性威胁类似机器人造成的认同威胁,人们会因担心未来机器人和人之间的界限模糊而产生集体焦虑。其次,群体灭绝在历史上并非新鲜事,当人们想象自然灾害可能使其群体灭绝,甚至只是象征性地让其群体的标识永久消失时,他们都会陷入群体焦虑(Wohl et al., 2010)。对于机器人的迅速发展,人们也存在一定的灭绝威胁担忧。例如,一些学者和专家认为,机器智能等彻底变革性技术给人类带来了前所未有的风险,可能导致人类灭绝或摧毁地球孕育智慧生命的潜力(Bostrom, 2002; Khasawneh, 2018; Li & Huang, 2020; Yudkowsky, 2008)。科幻小说中也常描述机器人对人类的残酷伤害,以及今年在 ChatGPT 等大语言模型开始涌现出可能的类人行为时,各种科技领袖呼吁暂停相关研究,以避免人类陷入万劫不复的境地(Kahn, 2023)。这些几乎都是机器人威胁导致集体焦虑的现实版本。更为直接的证据在于,经济不平等会产生集体焦虑,尤其是对于低收入人群而言(Jetten et al., 2017);而由基尼系数体现的国家不平等与人们对人工智能或机器人构成普遍失业威胁的看法呈显著正相关(Shoss & Ciarlante, 2022)。基于此,我们可推测,感知机器人威胁会增加集体焦虑。

与个体焦虑通常产生的负面人际后果不同,集体焦虑产生的人际后果可能是建设性的(即亲社会的),也可能是破坏性的(即反社会的)(Wohl et al.,

2012)。人类社会一直充满全球性危机,如流行病、地震、气候变化和恐怖主义(Kahn et al., 2022)。这些危机使人们产生了对整个群体未来的焦虑与担心。一方面,这种集体焦虑会促进群体间合作,如协作解决问题、投资公共设施或向其他群体提供帮助(Fuesting et al., 2019; Kung et al., 2018; Römpke et al., 2019)。而另一方面,这种群体焦虑也被认为会导致社会纽带的解体(Drury, 2018),反而加剧个人主义与自私行为。如,全球流行病加剧了群体间的竞争感知、歧视与不信任(Gordils et al., 2021)。自然灾害后人们表现出明显更少的信任行为(Fleming et al., 2014)。机器人威胁作为一种随技术进步而出现的新的集体焦虑来源,其对人类亲社会行为的影响尚未得到实证检验。

基于机器人威胁不同于其他威胁的独特特征,本研究认为机器人威胁引发的集体焦虑应会降低人类之间的亲社会行为,而非促进亲社会行为。首先,感知到的资源稀缺使人们对威胁产生了不同的关注点(Miao et al., 2023)。当人们认为资源足够应对威胁时,会产生跨群体间共享命运的感觉,并促进群体间的合作共同应对威胁(Cortland et al., 2017; Dovidio et al., 2004)。然而,当现有资源被认为不足以共同应对威胁时,这会破坏集体焦虑下的共同目的感(Bavel et al., 2020),并阻碍人际及群体间的互惠(Butz & Yogeewaran, 2011; Frankenberg et al., 2012; Gamez-Djokic & Waytz, 2020; Uenal et al., 2021)。机器人威胁则具备后者的特征。如,来自33个OECD国家的面板数据发现,机器人和人工智能导致了就业岗位不足,增加了失业率(Bordot, 2022)。因此,虽然机器人引发人们的集体焦虑——人们将机器人视为人类集体的共同威胁而担心人类群体的未来,但机器人抢夺资源的客观现实或主观信念使人们面对这种共同威胁和集体焦虑时,更倾向于寻找“替罪羊”,即降低亲社会倾向。其次,集体焦虑下人们合作与互惠的前提之一是将彼此视为单一的、共同的群体,即不同的社会群体同等地成为共同威胁的目标,从而产生“我们共同受苦”的感觉(Drury, 2018)。然而,在存在明显不平等的情况下,即使面对相同的威胁,人们也不愿将所有人视为一个单一群体,无论是优势群体还是弱势群体(Drury, 2018; Vezzali et al., 2017)。机器人的发展可能加剧社会不平等(Howard & Borenstein, 2018),例如,机器人可能存在种族与性别偏见(Howard & Borenstein, 2018),更可能先替代中低技能的岗位

(Goyal & Aneja, 2020),加剧经济不平等(Goyal & Aneja, 2020)。这使得因机器人产生的集体焦虑难以促使人们将彼此视为一个单一群体,相反,更可能引发敌意,降低亲社会性。

基于以上视角,感知机器人威胁导致的集体焦虑可能会对人际关系造成负面影响,导致亲社会倾向的降低。也就是说,机器人威胁使人们产生了更高的集体焦虑——对人类未来的担忧,但却并未将人类团结在一起,因此集体焦虑反而降低了人们的亲社会倾向。以往关于集体焦虑与人际影响的研究多以流行病、自然灾害、恐怖主义或外群体威胁作为集体焦虑的来源(如 Jetten et al., 2021; Wohl et al., 2011)。而随着技术的发展,机器人和人工智能可能对人类造成的现实和独特性威胁也受到越来越多的关注(Yogeewaran et al., 2016; Zlotowski et al., 2017),成为引发人们群体焦虑的来源之一。但这种威胁造成的集体焦虑所产生的后续人际影响却暂未得到证实。人们究竟会愈将人类视为整体内群体而更加团结,还是在集体焦虑下寻找替罪羊而更显隔阂?本研究聚焦此问题,试图探索感知机器人威胁经由集体焦虑对亲社会倾向的可能影响。

## 1.2 感知机器人威胁影响亲社会倾向有何边界条件

感知机器人威胁对亲社会行为的影响会因内外群体关系的不同而有所变化。根据群际威胁理论,在缺乏冲突的情况下,既不会出现强烈的内群体依恋,也不会出现外群体敌意(Brewer, 1999, p. 431; Sherif, 1966; Sumner, 1906)。而威胁是凸显和强化外群体划分的一个重要因素(Stephan et al., 2009),并会进一步引发对外群体的敌意(Riek et al., 2006)。即使是与外群体共同面对的威胁也会导致消极态度(Gordils et al., 2021; Miao et al., 2023)。例如,全球流行病增加了对白人与黑人间竞争的感知,以及随后的对外群体的歧视、行为回避、群体间焦虑和种族间不信任(Gordils et al., 2021);自然灾害后反移民情绪增加(Andrighetto et al., 2016);宏观经济威胁仅导致对亚裔美国人的偏见增加(Butz & Yogeewaran, 2011)。集体焦虑会增加对穆斯林的偏见(Tabri et al., 2018)、对移民的反对和偏见、不容忍民族间通婚(Jetten et al., 2021; Jetten & Wohl, 2012; Lucas et al., 2014; Roccas & Amit, 2011)等。因此,与其他威胁类似,机器人威胁可能强化人际关系中内外群体的界限与对立。一些研究提供了初步的证据。例如,对自动化的担忧增加人们对外

群体的反对和歧视 (Gamez-Djokic & Waytz, 2020)。自动化威胁使人们反对移民并支持限制移民和外国商品的政策 (Wu, 2022a, 2022b)。基于此, 我们假设内外群体的划分会调节感知机器人威胁对亲社会倾向的影响, 即感知机器人威胁降低亲社会倾向的效应对外群体成员更强, 而对内群体成员更弱。

内外群体的区分使个体在群体横向比较中调整对帮助他人的态度。作为一种与道德相关的行为, 机器人威胁对亲社会行为的影响可能会受到道德比较的影响。既往研究表明, 共同的外部威胁既可能促进合作与互助, 从而增加亲社会行为 (Fuesting et al., 2019; Kung et al., 2018; Römpke et al., 2019), 也可能助长自私行为, 降低亲社会性 (Fleming et al., 2014; Gordils et al., 2021)。这种双重效应可能源于人们在面对威胁时的动机差异: 一方面, 个体可能被鼓励超越自我关注, 考虑“他者”的处境; 另一方面, 个体可能通过正当性辩护来维护危机下自我和群体的地位。这涉及到向上还是向下的道德比较倾向。首先, 传统的现实群体冲突观点认为, 威胁会增加群体间的偏见与冲突, 而道德优越感可以为发生冲突时对外群体的支配或主动征服提供理由或

正当化 (Sidanius, 1993)。道德优越也为危机关头是谁有权得到这些资源, 谁应该得到这些资源, 以及这些资源如何分配提供了辩护 (Hirschberger & Pyszczynski, 2011)。因此, 下行道德比较产生的自己相对于他人的优越感可能导致人们在面对机器人威胁时表现得更自利。此外, 向下的道德比较可能使个体感到道德上的许可 (moral license; Lasarov & Hoffmann, 2020; Merritt et al., 2010), 从而允许他们在应对机器人威胁时放纵自己, 采取一些不道德的行为或减少亲社会行为以维护个人利益。例如, 在大屠杀诱导的威胁下, 道德许可增加了人们对伤害平民的接受度 (Hirschberger & Pyszczynski, 2011)。基于以上视角, 下行道德比较者在面对威胁时优先考虑自我而非他人的处境, 通过道德优越和道德许可为自己辩护。基于此, 我们假设感知机器人威胁对亲社会倾向的影响会受到道德比较方向的调节, 即感知机器人威胁降低亲社会倾向的效应在下行道德比较者中更强, 而在上行道德比较者中更弱。

### 1.3 研究概览

本文由 8 个递进研究组成 (表 1), 主要研究目的是探究感知机器人威胁对亲社会倾向的影响, 其

表 1 研究总结

研究	方法	实验设计	自变量测量/操纵	因变量	主要发现	样本
1a	档案数据库分析	—	感知机器人威胁量表 (the Eurobarometer 87.1, 2017)	捐赠指数 (the Eurobarometer 87.1, 2017)	感知机器人威胁与捐赠指数负相关。	27901
1b	档案数据库分析	—	感知机器人威胁量表 (the Eurobarometer 77.1, 2012)	对人道主义援助的支持 (the Eurobarometer 77.1, 2012)	感知机器人威胁与对人道主义援助的支持负相关。	26751
2	问卷调查	—	感知机器人威胁量表 (Yogeeswaran et al., 2016)	亲社会倾向 (Osgood & Muraven, 2015)	感知机器人威胁与亲社会倾向负相关, 集体焦虑在其中起中介作用。	148
3	线上实验	机器人威胁(高 vs. 低)	视频操纵 (Yogeeswaran et al., 2016; Zlotowski et al., 2017)	亲社会倾向 (Osgood & Muraven, 2015)	感知机器人威胁降低亲社会倾向, 集体焦虑在其中起中介作用。	270
4	线上实验	机器人威胁(高 vs. 低)	写作操纵 (许丽颖 等, 2024)	亲社会倾向 (Touré-Tillery & Light, 2018)	感知机器人威胁降低亲社会倾向, 集体焦虑在其中起中介作用。	270
5	线上实验	机器人威胁(高 vs. 低)×内外群体(内 vs. 外)	新闻网页图片操纵 (Jackson et al., 2020; 许丽颖 等, 2024)	亲社会倾向 (Osgood & Muraven, 2015)	感知机器人威胁降低对外群体的亲社会倾向, 但不会降低对内群体的亲社会倾向。	400
6	线上实验	机器人威胁(高 vs. 低)	新闻网页图片操纵 (Jackson et al., 2020; 许丽颖 等, 2024)	亲社会倾向 (Osgood & Muraven, 2015)	感知机器人威胁降低道德下行比较者的亲社会倾向, 但不会降低道德上行比较者的亲社会倾向。	270
7	行为实验	机器人威胁(高 vs. 低)	写作操纵 (许丽颖 等, 2024)	给需要帮助的人写鼓励信 (Gaesser et al., 2020)	感知机器人威胁降低亲社会行为。	161

中集体焦虑作为中介机制,而内外群体和道德比较倾向则会调节其效应。研究 1a 和 1b 用已有档案数据库发现了感知机器人威胁与亲社会行为之间的基本关系。研究 2 和研究 3 则分别以问卷研究和实验研究来探究感知机器人威胁如何影响亲社会倾向,即验证集体焦虑的中介机制。研究 4 排除了其他可能中介的影响。研究 5 和研究 6 分别探讨感知机器人威胁降低亲社会倾向的边界条件,即验证内外群体与道德比较倾向的调节作用。研究 7 探讨了感知机器人对真实亲社会行为的影响。

## 2 研究 1a: 感知机器人威胁与亲社会倾向(捐赠指数)的关系

研究 1a 从已有档案数据出发,初步确定感知

机器人威胁与亲社会行为之间的关系。在本研究中,以国家捐赠指数作为亲社会行为的指标。

### 2.1 样本

本研究的数据分为个体层面和群体层面。个体层面的数据来自欧洲晴雨表 87.1 (the Eurobarometer 87.1, 2017),其专题之一“数字化和自动化对日常生活的影响”涵盖了有关感知机器人威胁的内容。欧洲晴雨表 87.1 数据是 2017 年 3 月 18 日至 2017 年 3 月 27 日对具有代表性的欧洲公民(15 岁及以上)进行面对面访谈收集的,总样本包括来自 28 个欧盟成员国的 27 901 名被试的回复。国家层面的人口统计学信息见表 2。

在本研究中,我们将欧洲晴雨表 87.1 中的总计 27 901 名被试的回复全部纳入样本 1 的个体层面数

表 2 国家层面的人口统计信息: 性别和年龄

国家	样本量(N)		性别(女性占比)		年龄(M/SD)	
	全部样本	劳动力样本	全部样本	劳动力样本	全部样本	劳动力样本
1 法国	1004	412	56.37%	52.40%	52.13 (19.06)	42.19 (11.85)
2 比利时	1023	430	51.22%	50.00%	52.64 (19.01)	43.81 (11.98)
3 荷兰	1015	544	49.75%	48.70%	52.41 (16.39)	47.36 (12.12)
4 德国	1537	650	49.19%	50.70%	54.07 (19.39)	44.49 (12.22)
5 意大利	1022	471	54.01%	54.00%	49.3 (17.29)	45.72 (10.50)
6 卢森堡	510	221	52.75%	52.70%	50.87 (18.53)	43.13 (11.51)
7 丹麦	1000	495	50.30%	49.90%	55.58 (17.87)	48.52 (11.94)
8 爱尔兰	1021	509	51.91%	47.90%	48.60 (17.50)	43.62 (12.01)
9 英国	1346	620	50.45%	50.40%	53.27 (19.63)	43.40 (13.76)
10 希腊	1010	442	52.18%	45.50%	49.37 (17.97)	43.24 (12.06)
11 西班牙	1024	408	51.37%	45.60%	49.87 (18.43)	42.67 (11.49)
12 葡萄牙	1061	600	55.80%	55.20%	49.64 (18.01)	43.07 (11.92)
13 芬兰	1012	403	53.95%	52.30%	55.33 (18.89)	45.38 (12.77)
14 瑞典	1007	507	42.80%	40.80%	58.02 (17.24)	48.64 (13.52)
15 奥地利	1001	588	52.85%	50.90%	48.80 (16.96)	42.31 (11.58)
16 塞浦路斯	501	233	52.30%	46.40%	51.00 (18.16)	44.16 (11.70)
17 捷克共和国	1058	621	60.11%	55.90%	47.77 (16.64)	43.55 (11.58)
18 爱沙尼亚	1017	489	64.70%	60.90%	56.16 (18.31)	47.34 (12.27)
19 匈牙利	1053	525	56.98%	49.70%	52.76 (17.60)	43.55 (11.72)
20 拉脱维亚	1004	493	62.05%	61.30%	48.64 (17.35)	44.67 (13.62)
21 立陶宛	1001	404	64.24%	57.20%	55.66 (19.10)	45.00 (12.86)
22 马耳他	500	187	57.80%	48.70%	54.85 (19.00)	43.72 (14.35)
23 波兰	1008	514	59.23%	55.30%	48.74 (17.30)	42.08 (11.68)
24 斯洛伐克	1014	481	60.45%	56.80%	51.21 (17.53)	43.41 (11.40)
25 斯洛文尼亚	1027	438	59.69%	55.70%	52.95 (18.55)	43.89 (11.59)
26 保加利亚	1044	589	53.07%	52.10%	48.95 (16.82)	43.83 (11.14)
27 罗马利亚	1033	528	56.53%	49.20%	45.17 (16.58)	40.14 (11.02)
28 克罗地亚	1048	492	60.21%	58.30%	44.84 (17.11)	40.13 (11.05)
总计	27,901	13,294	54.97%	51.50%	51.38 (18.25)	44.06 (12.20)

据中(女性占比 55.0%,  $M_{\text{年龄}} = 51.38$  岁,  $SD = 18.25$  岁), 以期进行全面的和大规模的分析; 同时, 我们将样本 2 限制在 13 294 名在职被试(女性占比 51.5%,  $M_{\text{年龄}} = 44.06$  岁,  $SD = 12.20$  岁), 以期对劳动力样本进行单独分析, 因为以往研究在进行类似分析时认为, 机器人的发展对劳动力产生的威胁更大(Shoss & Ciarlante, 2022)。在样本 1 和样本 2 中(详见表 2), 我们都将欧洲晴雨表中西德和东德的样本合并作为德国样本(total  $N_1 = 1537$ , total  $N_2 = 650$ )分析, 将北爱尔兰和大不列颠的样本合并作为英国样本(total  $N_1 = 1346$ , total  $N_2 = 620$ )分析。表 2 包含样本 1 和样本 2 的国家层面的样本量以及各国家基本人口统计数据, 即性别、年龄和样本量。

## 2.2 测量与分析

### 2.2.1 个体层面的变量

**感知机器人威胁。**欧洲晴雨表中与感知机器人威胁相关的题目包括 6 道题: 1)“由于机器人和人工智能的使用, 更多的工作岗位将消失, 而不会创造更多的新工作岗位”; 2)“机器人和人工智能对社会来说是一件好事, 因为它们帮助人们完成工作或在家里完成日常任务”; 3)“机器人和人工智能是需要谨慎管理的技术”; 4)“机器人是必要的, 因为它们可以做对人类来说太困难或太危险的工作”; 5)“在工业数字化转型方面, 欧盟走在世界其他地区的前面”; 6)“机器人和人工智能抢走了人们的工作”(其中条目 2 和 4 为反向计分)。Shoss 和 Ciarlante (2022) 认为机器人主要对劳动力产生威胁, 因此他们将样本限制在在职人员, 并在以上条目中选取了与感知工作威胁最相关的两个条目(即条目 1 与条目 6)作为感知机器人威胁。我们参照 Shoss 和 Ciarlante (2022) 的做法将这两个条目的得分平均后作为感知机器人威胁指标 1。但我们认为机器人带来的威胁不局限于工作场所, 因此我们再选取条目 1、2、3、4 和 6 共 5 个条目, 将这 5 个条目得分平均后作为感知机器人威胁指标 2。以上条目均为李克特 4 点计分(1 = 完全不同意, 4 = 完全同意), 得分越高表明被试感知到机器人威胁程度更高。

**个体层面的控制变量。**本研究考虑了几个可能预测人们感知机器人威胁的个人层面的人口统计学和技术相关变量。人口统计变量包括教育、性别、年龄、社区规模和政治信念(McClure, 2018); 在技术相关变量上, 欧洲晴雨表调查了人们对先进技术的经验和知识, 包括被试的数字技术技能(即“掌握的这些技能是否足以支撑当前或未来的工作)、被

试对人工智能的知识(即, 是否在过去一年中阅读过或看到过有关人工智能的东西)以及在工作中使用机器人的情况(即, 他们目前是否使用或曾经在工作中使用机器人)。以往的研究表明, 这类变量也可能影响人们对机器人威胁的感知(Vu & Lim, 2021)。并且先前对欧洲晴雨表数据集的分析也发现, 这些变量与人们对技术影响的想法和接受度有关(Carradore, 2022; Rughiniş et al., 2018)。最后, 个体对于人工智能或机器人是否能在未来完成自己当前的工作的看法也受到了控制, 因为先进技术取代自己工作的威胁可能会影响个人对这项技术威胁性的一般看法(Shoss & Ciarlante, 2022)。纳入这些控制变量可以减少由于不同国家在这些因素上的系统性差异对结果所造成的影响。

### 2.2.2 国家层面的变量

**个体层面的数据(包括机器人威胁和其他控制变量)均来源于欧洲晴雨表 87.1, 鉴于该数据截至 2017 年 3 月收集, 于是国家层面的数据(包括亲社会行为与其他控制变量)采用 2016 年的指标更能与欧洲晴雨表 87.1 的数据在时间上保持一致。**

**亲社会行为。**亲社会的数据来源于世界慈善援助基金会(Charities Aid Foundation, CAF)的世界捐赠指数 2017 版(CAF World Giving Index 2017), 该版本反映的是 2016 年的世界捐赠指数。世界捐赠指数关注世界各地的人们的慈善捐赠事业, 包含三种不同类型的捐赠行为: 帮助陌生人(是否帮助陌生人或不认识但需要帮助的人?), 金钱捐赠(是否向慈善机构捐赠金钱?)和时间付出(是否志愿向某个组织服务?)。这三种不同的捐赠行为分别形成了帮助陌生人指数、金钱捐赠指数和时间付出指数, 将这三个指数平均后得到捐赠指数, 得分越高表明亲社会行为越多。该报告收集了来自 139 个国家的数据, 我们从中采用了与欧洲晴雨表 87.1 对应的 28 个欧盟国家的数据进行分析。

**国家层面的控制变量。**由于亲社会行为可能受到经济发展水平的影响, 我们控制了 2016 年各国的人均国内生产总值(gross domestic product per capita, GDP per capita), 即一个国家的经济产出除以其人口的年度估计值。这一数据来自世界发展指标(World Bank, 2016)。为了确保数据的正态, 我们取人均国内生产总值的对数(lnGDP), 并在所有分析中使用这些对数转换后的值。由于亲社会行为也与公平相关, 我们还控制了 2016 年的国家基尼系数(Gini index)。本研究国家层面的经济指标来自世

界银行(the World Bank)的世界发展指标(World Development Indicators, WDI; World Bank, 2016)。国家层面的亲社会行为数据与控制变量指数见表 3。

2.2.3 分析方法

本研究以亲社会倾向(捐赠指数)为因变量, 感知机器人威胁为自变量, 性别、年龄、社区、教育、政治取向、技术技能、对机器人的相关了解、机器人使用、社会阶层、国家 Gini 系数和国家 GDP 为控制变量, 使用 SPSS 25 进行层次线性回归分析。回归公式如下:

$$y_{j(i)} = \beta_1 x_i + \text{Control}_i + \text{Control}_j + \varepsilon_i$$

其中  $i$  表示个体,  $j$  表示国家,  $j(i)$  表示个体  $i$  所处的国家,  $y_{j(i)}$  是 2016 年的捐赠指数,  $x_i$  是机器人

威胁指标,  $\text{Control}_i$  表示个体层面的控制变量,  $\text{Control}_j$  表示国家层面的控制变量,  $\varepsilon_i$  表示随机误差。

2.3 结果

2.3.1 所有样本

表 4 提供了所有样本描述性统计数据和相关性。从表 4 中可以看出, 由 2 条目合成的机器人威胁指标 1 与帮助他人指数( $r = -0.065, p < 0.001$ )、金钱捐赠指数( $r = -0.151, p < 0.001$ )、时间付出指数( $r = -0.083, p < 0.001$ )以及总的捐赠指数( $r = -0.125, p < 0.001$ )都显著负相关。由 5 条目合成的机器人威胁指标 2 也与帮助他人指数( $r = -0.014, p = 0.017$ )、金钱捐赠指数( $r = -0.102, p < 0.001$ )、时间付出指数( $r = -0.017, p < 0.001$ )以及总的捐赠指数( $r = -0.063, p < 0.001$ )都显著负相关。

表 3 亲社会行为指标与国家层面的控制变量

国家	亲社会行为指数(%)				ln (GDP)	Gini 指数	
	捐赠指数	帮助他人	金钱捐赠	时间付出			
1	法国	33	39	30	31	10.52	31.90
2	比利时	35	46	34	26	10.65	27.60
3	荷兰	51	51	64	36	10.74	28.20
4	德国	45	58	55	22	10.65	31.40
5	意大利	30	44	30	15	10.34	35.20
6	卢森堡	38	37	48	28	11.58	31.70
7	丹麦	44	57	54	21	10.91	28.20
8	爱尔兰	53	61	60	39	11.05	32.80
9	英国	50	58	64	28	10.62	33.10
10	希腊	24	50	10	11	9.79	35.00
11	西班牙	33	51	33	14	10.19	35.80
12	葡萄牙	26	46	14	17	9.90	35.20
13	芬兰	40	55	37	28	10.69	27.10
14	瑞典	41	53	55	14	10.86	29.60
15	奥地利	42	51	48	28	10.72	30.80
16	塞浦路斯	38	54	34	24	10.12	32.90
17	捷克共和国	18	23	18	14	9.83	25.40
18	爱沙尼亚	27	36	22	22	9.81	31.60
19	匈牙利	21	36	17	9	9.48	30.30
20	拉脱维亚	18	28	20	5	9.57	34.30
21	立陶宛	16	28	10	10	9.62	38.40
22	马耳他	48	45	73	26	10.15	29.10
23	波兰	26	37	27	13	9.42	31.20
24	斯洛伐克	26	33	30	16	9.71	25.20
25	斯洛文尼亚	34	40	32	32	9.98	24.80
26	保加利亚	19	34	17	5	8.93	40.60
27	罗马利亚	31	60	24	9	9.15	34.40
28	克罗地亚	20	21	28	12	9.44	30.90



续表

变量	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
18 机器人威胁 <sup>b</sup>	-0.053***	0.073***	-0.030***	-0.127***	-0.015*	-0.226***	-0.203***	-0.204***	-0.123***	0.146***	0.111***	0.070***	0.024***	-0.092***	0.01	0.034***	0.102***	—	—	—	—	—	—	—	—	—	—	
19 机器人威胁 <sup>c</sup>	0.01	0.030***	0.00	0.080***	-0.018**	0.101***	0.132***	0.094***	-0.033***	-0.155***	0.01	0.00	0.029***	0.00	0.025***	0.075***	0.216***	-0.213***	—	—	—	—	—	—	—	—	—	
20 机器人威胁 <sup>d</sup>	-0.051***	0.00	-0.018**	-0.116***	-0.01	-0.179***	-0.192***	-0.164***	-0.052***	0.184***	0.059***	0.047***	0.00	-0.052***	0.00	-0.012*	-0.017**	0.472***	-0.376***	—	—	—	—	—	—	—	—	
21 机器人威胁 <sup>e</sup>	-0.069**	0.050***	-0.058***	-0.185***	-0.01	-0.162***	-0.157***	-0.141***	0.045***	0.169***	0.054**	0.070**	0.030**	-0.189**	0.171**	-0.126**	0.638**	0.159**	0.158**	0.044**	—	—	—	—	—	—	—	
22 机器人威胁指标 <sup>1</sup>	0.041**	-0.058***	-0.177***	-0.020**	-0.152***	-0.141***	-0.127***	0.046***	0.157***	0.056**	0.067**	0.036**	-0.186**	0.167**	-0.119**	0.902***	0.145***	0.202***	0.017**	0.912***	—	—	—	—	—	—	—	
23 机器人威胁指标 <sup>2</sup>	0.069***	-0.057***	-0.188***	-0.026**	-0.222***	-0.203***	-0.193***	-0.041**	0.182***	0.100**	0.087**	0.045**	-0.179**	0.130**	-0.053**	0.719***	0.578***	0.238***	0.423***	0.750***	0.807***	—	—	—	—	—	—	
24 帮助他人	0.076**	0.045***	-0.014*	0.068**	-0.067**	0.097**	0.108**	0.082**	-0.043**	-0.086**	0.00	-0.027**	-0.017**	0.066**	0.020**	0.580***	-0.050**	0.028**	0.045***	0.017**	-0.071**	-0.065**	-0.014*	—	—	—	—	
25 金钱捐赠	0.064***	0.063***	-0.057**	0.119***	-0.073**	0.162***	0.198***	0.164***	-0.093**	-0.176**	-0.025**	-0.035**	0.00	0.150***	-0.320**	0.759***	-0.115**	-0.01	0.051***	-0.027**	-0.159**	-0.151**	-0.102**	0.625***	—	—	—	
26 时间付出	0.031**	0.053**	-0.070**	0.050**	-0.058**	0.106**	0.136**	0.129**	-0.090**	-0.162**	-0.014*	-0.039**	-0.01	0.114**	-0.445**	0.755***	-0.057**	0.062**	0.036**	0.023**	-0.094**	-0.083**	-0.017**	0.473**	0.681**	—	—	
27 捐赠指数	0.070**	0.065***	-0.053**	0.101**	-0.077**	0.150**	0.180**	0.151**	-0.089**	-0.169**	-0.017**	-0.038**	-0.01	0.134**	-0.272**	0.812***	-0.094**	0.023**	0.055**	0.00	-0.134**	-0.125**	-0.063**	0.817**	0.935***	0.795**	—	
M	51.38	1.95	19.66	5.28	2.18	1.07	2.89	3.18	2.99	1.71	0.53	0.93	0.95	0.98	2.35	31.56	10.14	3.16	2.09	3.49	1.70	3.09	3.12	2.71	33.00	44.29	34.98	19.57
SD	18.25	0.77	5.23	2.18	1.07	0.92	0.96	0.88	0.50	0.26	0.26	0.22	0.13	0.99	3.83	0.61	0.85	0.84	0.67	0.77	0.89	0.79	0.47	11.00	11.59	17.38	9.25	

注: 技术技能 a = 掌握的这些技能是否足以支撑当前工作; 技术技能 b = 掌握的这些技能是否足以支撑未来工作; 工作被替代 = 当前工作中是否可能被机器人取代; 相关了解 = 在过去 12 个月是否阅读过或使用过机器人; 机器人使用 a = 是否在家使用过机器人; 机器人使用 b = 是否在工作中使用过机器人; 机器人使用 c = 是否在其他地方使用过机器人; 机器人威胁 a = “由于机器人的使用, 更多的工作岗位将消失, 而不会创造更多的新工作岗位”; 机器人威胁 b = “机器人和人工智能对社会来说是一件好事, 因为它们帮助人们完成工作或在家里完成日常工作”; 机器人威胁 c = “机器人和人工智能是需要谨慎管理的技术”; 机器人威胁 d = “机器人是必要的, 因为它们可以对人类来说太困难或太危险的工作”; 机器人威胁 e = “机器人和人工智能抢走了人们的工作”; 机器人威胁指标 1 = 机器人威胁 a 与机器人威胁 c 两条目平均分; 机器人威胁指标 2 = 机器人威胁 a, b, c, d, e 五条目平均分; 捐赠指数 = 帮助他人、金钱捐赠和时间付出三个指数的平均分。  
\* $p < 0.05$ . \*\* $p < 0.01$ . \*\*\* $p < 0.001$ .

由于许多控制变量与捐赠指数相关,因此我们以捐赠指数为因变量,2 条目合成的机器人威胁指标 1 为自变量,并将控制变量纳入模型进行分层线性回归分析。结果发现,在排除了人口统计相关的控制变量(即性别、年龄、社区、教育和政治取向)、技术相关控制变量(即技术技能、相关了解和机器人使用)以及国家层面相关控制变量(Gini 系数和国家 GDP)的影响后,机器人威胁指标 1 仍负向影响亲社会行为,  $B = -0.51, p < 0.001$ , 详见表 5。

表 5 全部样本中的机器人威胁与亲社会行为

变量	回归系数与显著性	变量	回归系数与显著性
(常量)	-121.36(1.111)	(常量)	-121.70(1.119)
性别	0.38*** (0.090)	性别	0.38*** (0.090)
年龄	0.01*** (0.003)	年龄	0.01** (0.003)
社区	-0.35*** (0.058)	社区	-0.34*** (0.058)
教育	-0.04*** (0.009)	教育	-0.04*** (0.009)
政治取向	-0.11*** (0.021)	政治取向	-0.11*** (0.021)
技术技能 a	0.43*** (0.052)	技术技能 a	0.41*** (0.052)
相关了解	0.29** (0.098)	相关了解	0.29** (0.098)
机器人使用 a	1.35*** (0.167)	机器人使用 a	1.40*** (0.167)
机器人使用 c	0.28(0.338)	机器人使用 c	0.31(0.338)
社会阶层	-0.01(0.050)	社会阶层	0.00(0.050)
Gini 系数	0.06*** (0.013)	Gini 系数	0.06*** (0.013)
ln GDP	15.03*** (0.081)	ln GDP	15.07*** (0.080)
机器人威胁指标 1	-0.51*** (0.059)	机器人威胁指标 2	-0.60*** (0.102)

注: 技术技能 a = 掌握的这些技能是否足以支撑日常生活; 相关了解 = 在过去 12 个月是否阅读过或接触过与人工智能或机器人有关的信息; 机器人使用 a = 是否在家使用过机器人; 机器人使用 c = 是否在其他地方使用过机器人; 机器人威胁指标 1 = 机器人威胁 a 与机器人威胁 e 两条目平均得分; 机器人威胁指标 2 = 机器人威胁 a, b, c, d, e 五条目平均得分。

\*\* $p < 0.01$ . \*\*\* $p < 0.001$ . 括号内为标准误。

同样,以捐赠指数为因变量,5 条目合成的机器人威胁指标 2 为自变量,并将控制变量纳入模型进行分层线性回归分析发现,在排除了人口统计相关的控制变量(即性别、年龄、社区、教育和政治取向)、技术相关控制变量(即技术技能、相关了解和机器人使用)和国家层面相关控制变量(Gini 系数和国家 GDP)的影响后,机器人威胁指标 2 仍负向影响亲社会行为,  $B = -0.60, p < 0.001$ , 详见表 5。

### 2.3.2 劳动力样本

表 6 提供了劳动力样本的描述性统计数据和相关性。从表 6 中可以看出,由 2 条目合成的机器人威胁指标 1 与帮助他人指数( $r = -0.069, p < 0.001$ )、

金钱捐赠指数( $r = -0.172, p < 0.001$ )、时间付出指数( $r = -0.098, p < 0.001$ )以及总的捐赠指数( $r = -0.141, p < 0.001$ )都显著负相关。而由 5 条目合成的机器人威胁指标 2 只与金钱捐赠指数( $r = -0.096, p < 0.001$ )和总的捐赠指数( $r = -0.057, p < 0.001$ )显著负相关。

同样,我们以捐赠指数为因变量,2 条目合成的机器人威胁指标 1 为自变量,并将控制变量纳入模型进行分层线性回归分析。结果发现,在排除了人口统计相关的控制变量(即性别、社区、政治取向与社会阶层)、技术相关控制变量(即技术技能、相关了解和机器人使用)和国家层面相关控制变量(Gini 系数和国家 GDP)的影响后,机器人威胁指标 1 仍负向影响亲社会行为,  $B = -0.60, p < 0.001$ , 详见表 7。

同样,以捐赠指数为因变量,5 条目合成的机器人威胁指标 2 为自变量,并将控制变量纳入模型进行分层线性回归分析发现,在排除了人口统计相关的控制变量(即性别、社区、政治取向与社会阶层)、技术相关控制变量(即技术技能、工作被替代性和机器人使用)和国家层面相关控制变量(Gini 系数和国家 GDP)的影响后,机器人威胁指标 2 仍负向影响亲社会行为,  $B = -0.63, p < 0.001$ , 详见表 7。

## 2.4 讨论

研究 1a 通过档案数据库回溯的方法初步发现了感知机器人威胁和亲社会行为的负向关系。然而,研究 1a 存在一些局限。使用国家层面的捐赠指数与个体层面的感知机器人威胁进行分析可能忽略了个体间的差异,使得结果存在偏差或造成区群谬误。此外,亲社会行为使用的国家层面的数据(捐赠指数),这可能导致因变量受到国家经济、风俗和观念的较大影响。因此,研究 1b 将使用个体层面数据进行档案数据库回溯分析,并控制国家固定效应。

## 3 研究 1b: 感知机器人威胁与亲社会倾向(对人道主义援助的支持)的关系

为了提高档案分析研究结果的稳健性,研究 1b 在研究 1a 的基础上进行了改进。首先,研究 1b 采用了个体层面的感知机器人威胁和亲社会倾向进行分析,以减少由研究 1a 中跨层次分析所引起的结果偏差。其次,研究 1b 以对人道主义援助的支持作为亲社会倾向的指标,旨在增加因变量的多样性。第三,研究 1b 控制了国家固定效应,以减少国



续表

变量	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
18 机器人威胁 <sup>b</sup>	-0.032***	0.01	-0.028**	-0.101***	-0.020*	-0.205***	-0.203***	-0.195***	-0.123***	0.127***	0.116***	0.073***	0.02	-0.074**	-0.018*	0.052***	0.094**	—	—	—	—	—	—	—	—	—	—	—
19 机器人威胁 <sup>c</sup>	0.00	0.051***	0.00	0.072***	-0.01	0.149***	0.132***	0.101***	-0.033***	-0.159***	0.01	0.01	0.036***	-0.01	0.036***	0.077***	0.199***	-0.204**	—	—	—	—	—	—	—	—	—	—
20 机器人威胁 <sup>d</sup>	-0.033***	-0.034***	-0.022*	-0.094***	-0.020*	-0.198***	-0.192***	-0.160***	-0.052***	0.169***	0.065***	0.051***	-0.01	-0.038***	-0.024**	0.00	-0.02	0.458***	-0.365***	—	—	—	—	—	—	—	—	—
21 机器人威胁 <sup>e</sup>	-0.048***	0.00	-0.065***	-0.191***	0.00	-0.124***	-0.157***	-0.136***	0.045***	0.165***	0.049***	0.068***	0.01	-0.192***	0.187***	-0.141***	0.626***	0.140***	0.135***	0.042***	—	—	—	—	—	—	—	—
22 机器人威胁指标 <sup>1</sup>	-0.056***	-0.01	-0.063***	-0.178***	-0.01	-0.112***	-0.141***	-0.125***	0.046***	0.150***	0.053***	0.067***	0.018*	-0.185***	0.189***	-0.132***	0.898***	0.129***	0.180***	0.01	0.909***	—	—	—	—	—	—	—
23 机器人威胁指标 <sup>2</sup>	-0.064***	0.00	-0.064***	-0.176***	-0.023*	-0.181***	-0.203***	-0.187***	-0.041**	0.166***	0.106***	0.094***	0.027**	-0.174***	0.132***	-0.049***	0.723***	0.562***	0.231***	0.415***	0.748***	0.810***	—	—	—	—	—	—
24 帮助他人	0.076**	0.032***	-0.025**	0.106***	-0.119***	0.099***	0.108***	0.118***	-0.043***	-0.090***	0.00	-0.023**	-0.020*	0.049***	0.041**	0.580***	-0.049***	0.036***	0.060***	0.019*	-0.078***	-0.069***	-0.01	—	—	—	—	—
25 金钱捐赠	0.048***	0.055***	-0.050***	0.149***	-0.122***	0.166***	0.198***	0.192***	-0.093***	-0.177***	-0.026**	-0.041**	0.00	0.157***	-0.315***	0.775***	-0.126***	0.019*	0.061***	0.00	-0.185***	-0.172***	-0.096***	0.632***	—	—	—	—
26 时间付出	0.019*	0.033***	-0.079***	0.066***	-0.115***	0.112***	0.136***	0.135***	-0.090***	-0.150***	-0.022**	-0.047***	-0.01	0.112***	-0.429***	0.766***	-0.066***	0.081***	0.036***	0.038***	-0.109***	-0.098***	-0.02	0.486***	0.698***	—	—	—
27 捐赠指数	0.057***	0.051***	-0.056***	0.134***	-0.135***	0.154***	0.180***	0.179***	-0.089***	-0.167***	-0.019*	-0.041***	-0.01	0.130***	-0.256***	0.818***	-0.101***	0.043***	0.066***	0.01	-0.154***	-0.141***	-0.057***	0.820***	0.936***	0.805***	—	—
M	44.06	1.98	20.55	5.38	3.23	3.18	3.03	3.03	1.71	0.47	0.91	0.92	0.98	2.47	31.57	10.13	3.11	2.04	3.50	1.67	3.02	3.06	2.67	44.10	34.73	19.37	32.79	—
SD	12.20	0.79	4.75	2.15	0.85	0.92	0.93	0.88	0.50	0.29	0.29	0.27	0.14	0.97	3.86	0.62	0.85	0.81	0.65	0.75	0.89	0.79	0.45	11.71	17.42	9.35	11.13	—

注：技术技能 a = 掌握的这些技能是否足以支撑当前工作；技术技能 b = 掌握的这些技能是否足以支撑未来工作；工作被替代 = 当前工作在未来是否可能被机器人取代；相关了解 = 在过去 12 个月是否阅读过或接触过与人工智能或机器人有关的信息；机器人使用 a = 是否在家使用过机器人；机器人使用 b = 是否在工作中使用过机器人；机器人使用 c = 是否在其他地方使用过机器人；机器人威胁 a = “由于机器人和人工智能的使用，更多的工作岗位将消失，而不会创造更多的新工作岗位”；机器人威胁 b = “机器人和人工智能对社会来说是一件好事，因为它们帮助人们完成工作或在家里完成日常工作”；机器人威胁 c = “机器人和人工智能是需要谨慎管理的技术”；机器人威胁 d = “机器人是必要的，因为它们可以做对人类来说太困难或太危险的工作”；机器人威胁 e = “机器人和人工智能抢走了人们的工作”；机器人威胁指标 1 = 机器人威胁 a 与机器人威胁 c 两条目平均得分；机器人威胁指标 2 = 机器人威胁 b、c、d、e 五条目平均得分；捐赠指数 = 帮助他人、金钱捐赠和时间付出三个指数的平均得分。

\*  $p < 0.05$ . \*\*  $p < 0.01$ . \*\*\*  $p < 0.001$ .

表7 劳动力样本中的机器人威胁与亲社会行为

变量	回归系数与显著性	变量	回归系数与显著性
(常量)	-125.47(1.636)	(常量)	-125.90(1.654)
性别	0.38**(0.130)	性别	0.39**(0.130)
年龄	0.01(0.006)	年龄	0.01(0.006)
社区	-0.28**(0.083)	社区	-0.27**(0.083)
教育	-0.02(0.014)	教育	-0.01(0.014)
政治取向	-0.15*** (0.030)	政治取向	-0.15*** (0.030)
技术技能 a	-0.40** (0.127)	技术技能 a	-0.42** (0.127)
技术技能 b	0.54*** (0.120)	技术技能 b	0.54*** (0.120)
技术技能 c	0.58*** (0.108)	技术技能 c	0.58*** (0.108)
工作被替代	-0.13(0.075)	工作被替代	-0.17*(0.075)
相关了解	0.31*(0.141)	相关了解	0.29*(0.141)
机器人使用 a	1.03*** (0.223)	机器人使用 a	1.08*** (0.223)
机器人使用 b	0.48*(0.232)	机器人使用 b	0.46*(0.232)
机器人使用 c	0.26(0.456)	机器人使用 c	0.27(0.457)
社会阶层	-0.22** (0.073)	社会阶层	-0.19** (0.073)
Gini 系数	0.13*** (0.019)	Gini 系数	0.13*** (0.018)
ln GDP	15.19*** (0.116)	ln GDP	15.23*** (0.116)
机器人威胁指标 1	-0.60*** (0.085)	机器人威胁指标 2	-0.63*** (0.152)

注: 技术技能 a = 掌握的这些技能是否足以支撑日常生活; 技术技能 b = 掌握的这些技能是否足以支撑当前工作; 技术技能 c = 掌握的这些技能是否足以支撑未来工作; 工作被替代 = 当前工作在未来是否可能被机器人取代; 相关了解 = 在过去 12 个月是否阅读过或接触过与人工智能或机器人有关的信息; 机器人使用 a = 是否在家使用过机器人; 机器人使用 b = 是否在工作中使用过机器人; 机器人使用 c = 是否在其他地方使用过机器人; 机器人威胁指标 1 = 机器人威胁 a 与机器人威胁 c 两条目平均得分; 机器人威胁指标 2 = 机器人威胁 a, b, c, d, e 五条目平均得分。

\* $p < 0.05$ . \*\* $p < 0.01$ . \*\*\* $p < 0.001$ . 括号内为标准误。

家间不可观测因素(如风俗、观念)对结果的影响。最后, 研究 1b 采用了与 1a 不同时间点的数据, 以验证研究结果在不同时间的稳定性。

### 3.1 样本

本研究的数据来自欧洲晴雨表 77.1 (the Eurobarometer 77.1, 2012), 其专题“公众对机器人的态度”涵盖了有关感知机器人威胁的内容; 专题“人道主义援助”涵盖了有关亲社会倾向的内容。欧洲晴雨表 77.1 数据是 2012 年 2 月 25 日至 2012 年 3 月 12 日对具有代表性的欧洲公民(15 岁及以上)进行面对面访谈收集的, 总样本包括来自 27 个欧盟成员国的 26751 名被试的回复。

与研究 1a 一样, 我们将欧洲晴雨表中西德和东德的样本合并作为德国样本分析, 将北爱尔兰和大不列颠的样本合并作为英国样本分析。表 8 包含国家层面的样本量以及各国家基本人口统计数据, 即性别、年龄和样本量。

### 3.2 测量

#### 3.2.1 感知机器人威胁

欧洲晴雨表 77.1 中与感知机器人威胁相关的

题目包括 5 道题: (1)“机器人对社会是件好事, 因为它们帮助人们”; (2)“机器人抢走了人们的工作”; (3)“机器人是必要的, 因为它们可以做对人类来说太困难或太危险的工作”; (4)“机器人是一种需要谨慎管理的技术”; (5)“机器人的广泛应用可以增加欧盟的工作机会”。我们将 5 个条目得分平均后作为感知机器人威胁的分数。以上条目均为李克特 4 点计分(1 = 完全不同意, 4 = 完全同意), 其中, 条目 1、3 和 5 为反向计分。得分越高表明被试感知到机器人威胁程度更高。

#### 3.2.2 亲社会倾向

欧洲晴雨表 77.1 中与亲社会倾向相关的题目包括 4 道题: “人道主义援助是向欧盟以外国家在危机和紧急情况下需要帮助的人提供的援助。主要目标是挽救生命和减轻受影响人民的痛苦: (1)你认为欧盟资助这些人道主义援助活动有多重要? ; (2)欧洲正处于经济危机之中, 公共财政面临着巨大压力。尽管如此, 欧盟仍然继续资助人道主义援助, 以向世界各地急需援助的人们提供帮助。你对此有多支持? ; (3)你对欧盟的人道主义援助活动有多了

表 8 各国家的人口统计学信息和经济水平

	国家	样本量(N)	性别 (女性占比)	年龄(M/SD)	ln (GDP) (10 亿)
1	法国	1059	55.15%	49.97 (19.19)	3.46
2	比利时	1051	53.00%	49.81 (17.80)	2.72
3	荷兰	1014	52.86%	50.68 (18.61)	2.96
4	德国	1552	49.74%	52.44 (18.22)	3.57
5	意大利	1036	56.56%	47.44 (16.82)	3.36
6	卢森堡	501	54.69%	47.81 (17.87)	1.79
7	丹麦	1019	50.34%	50.84 (18.26)	2.54
8	爱尔兰	1008	54.27%	45.64 (17.14)	2.38
9	英国	1331	53.57%	50.74 (20.21)	3.42
10	希腊	999	53.25%	46.08 (17.36)	2.45
11	西班牙	1004	53.09%	47.18 (18.87)	3.17
12	葡萄牙	1009	54.51%	48.90 (18.74)	2.39
13	芬兰	1003	53.64%	55.58 (18.60)	2.44
14	瑞典	1016	48.33%	57.51 (15.78)	2.76
15	奥地利	1031	56.06%	46.01 (15.94)	2.64
16	塞浦路斯	506	50.99%	43.38 (17.29)	1.44
17	捷克共和国	1003	56.73%	46.12 (16.40)	2.36
18	爱沙尼亚	1000	62.60%	49.84 (19.26)	1.37
19	匈牙利	1021	56.90%	47.83 (17.37)	2.15
20	拉脱维亚	1024	53.52%	42.76 (16.86)	1.44
21	立陶宛	1021	54.95%	46.32 (18.12)	1.64
22	马耳他	500	62.00%	51.63 (17.88)	0.98
23	波兰	1000	58.90%	49.26 (18.85)	2.72
24	斯洛伐克	1000	58.20%	45.96 (15.32)	2.00
25	斯洛文尼亚	1017	52.02%	48.33 (17.90)	1.71
26	保加利亚	1006	51.99%	48.16 (16.82)	1.76
27	罗马尼亚	1020	47.65%	44.98 (16.61)	2.28
	总计	26, 751	54.08%	48.72 (18.08)	

解? ; (4)欧洲委员会正在建立一个‘欧洲自愿人道援助队’。来自欧盟国家的志愿者将被选拔、培训并部署到世界其他地区进行人道主义援助活动。你对此有多支持?”以上条目均为李克特 4 点计分。将 4 条目得分平均后得到亲社会指数,得分越高表明被试亲社会倾向更高。

### 3.2.3 个体层面的控制变量

本研究考虑了几个可能预测人们感知机器人威胁的个人层面的人口统计学和技术相关变量。人口统计变量包括性别、年龄、教育、社区规模、政治信念和社会阶层(McClure, 2018); 在技术相关变量上, 欧洲晴雨表调查了人们对先进技术的经验和知识, 包括被试对科学技术的兴趣(即“你对科学发现和技术发展的兴趣如何”）、使用机器人的情况(即“你是否曾经在家里或工作中使用过机器人, 如家用的扫地机器人或工作中的工业机器人”), 以及被

试对机器人的一般态度(即“一般来说, 你对机器人的态度如何?”)。以往的研究表明, 这类变量也可能影响人们对机器人威胁的感知(Carradore, 2022; Rughiniş et al., 2018; Vu & Lim, 2021)。以上个体层面的控制变量的数据均来自欧洲晴雨表 77.1。纳入这些控制变量可以减少由于不同国家在这些因素上的系统性差异对结果所造成的影响。

### 3.2.4 国家层面的控制变量

由于亲社会行为可能受到经济发展水平的影响, 我们控制了 2011 年各国的人均国内生产总值(gross domestic product per capita, GDP per capita), 即一个国家的经济产出除以其人口的年度估计值。这一数据来自世界发展指标(World Bank, 2011)。为了确保数据的正态, 我们取人均国内生产总值的对数(lnGDP), 并在所有分析中使用这些对数转换后的值(见表 8)。

### 3.2.5 分析方法

本研究以亲社会倾向(对人道主义援助支持)为因变量,感知机器人威胁为自变量,性别、年龄、社区、教育、社会阶层、科学兴趣、机器人使用、对机器人的一般态度和国家 GDP 为控制变量。使用 STATA 进行 HDFE (High dimensional fixed effects)线性回归分析,控制国家固定效应。回归公式如下:

$$y_i = \beta_1 x_i + \text{Control}_i + \text{Control}_j + \alpha_j + \varepsilon_i$$

其中  $i$  表示个体,  $j$  表示国家,  $y_i$  是亲社会指标,  $x_i$  是机器人威胁指标,  $\text{Control}_i$  表示个体层面的控制变量,  $\text{Control}_j$  表示国家层面的控制变量,  $\alpha_j$  表示国家固定效应,  $\varepsilon_i$  表示随机误差。

### 3.3 结果

表 9 提供了描述性统计数据和相关性。从表 9 中可以看出,感知机器人威胁与亲社会倾向负相关 ( $r = -0.154, p < 0.001$ )。

由于许多控制变量与感知机器人威胁或亲社会倾向相关,以及国家固定效应也可能影响二者。我们以亲社会倾向为因变量,感知机器人威胁为自变量,并将控制变量与国家固定效应纳入模型进行 HDFE 线性回归分析。

结果发现,在排除了人口统计相关的控制变量(即性别、年龄、社区、教育和社会阶层)、技术相关控制变量(即对科学技术的兴趣、机器人使用和

对机器人的一般态度)、国家层面相关控制变量(国家 GDP),以及国家固定效应的影响后,感知机器人威胁仍负向影响亲社会倾向,  $B = -0.077, p < 0.001$ , 详见表 10。

### 3.4 讨论

研究 1b 使用档案回溯的方法再次支持了感知机器人威胁和亲社会行为的负向关系。但与大多数档案研究一样,具有一些局限性(Barnes et al., 2018)。首先,档案分析不能揭示变量之间的因果关系。第二,用以档案分析的数据易受其他干扰因素的影响,如经济环境等。第三,包含感知机器人威胁与亲社会倾向相关的个体数据有限,如果有更近期的数据,研究结果也可能存在差异(Yam et al., 2023)。第四,我们采用的数据库仅限于单一时间点的采集。然而,若能够获取包含更多时间段的数据,将有助于提高研究结果在时间维度上的稳健性。

## 4 研究 2: 集体焦虑的中介作用(相关)

研究 2 以问卷调查的方式(截面研究)探讨感知机器人威胁与亲社会倾向的关系及集体焦虑在其中的中介作用。

### 4.1 被试

本研究采用 G\*Power 3.1 软件(Faul et al., 2007)确定所需样本量,对于本实验适用的相关分析,参

表 9 相关性描述性数据

变量	1	2	3	4	5	6	7	8	9	10	11	12	13
1 性别	—												
2 年龄	0.02***	—											
3 社区	0.01	-0.05***	—										
4 教育	-0.01*	-0.13***	0.13***	—									
5 社会阶层	-0.02***	-0.07***	0.04***	0.23***	—								
6 科学兴趣	-0.02***	-0.15***	0.08***	0.27***	0.19***	—							
7 机器人使用 <sup>a</sup>	0.01	-0.05***	0.01	0.03***	0.04***	0.06***	—						
8 机器人使用 <sup>b</sup>	-0.12***	-0.06***	-0.02**	0.06***	0.04***	0.12***	0.09***	—					
9 机器人使用 <sup>c</sup>	-0.02***	-0.04***	-0.02**	0.01*	0.02**	0.04***	0.02***	0.00	—				
10 机器人一般态度	-0.13***	-0.12***	0.05***	0.2***	0.13***	0.40***	0.12***	0.13***	0.05***	—			
11 ln GDP	-0.02***	0.07	0.01	-0.01*	0.10***	0.07***	0.02***	0.04***	-0.00	-0.01	—		
12 机器人威胁	-0.10**	0.06***	-0.04***	-0.20***	-0.14***	-0.26***	-0.08***	-0.10***	-0.04***	-0.55***	0.01	—	
13 亲社会倾向	-0.00	-0.05***	0.06***	0.15***	-0.11***	0.22***	0.03***	0.02**	0.00	-0.2***	-0.02**	-0.15***	—
<i>M</i>		48.72	1.92	19.02	5.54	2.01	0.05	0.06	0.01	2.87	2.46	2.52	2.99
<i>SD</i>		18.08	0.80	4.74	1.63	0.68	0.21	0.23	0.11	0.74	0.68	0.48	0.51

注: 机器人使用 a = 是否在家使用过机器人; 机器人使用 b = 是否在工作中使用过机器人; 机器人使用 c = 是否在其他地方使用过机器人。

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

表 10 感知机器人威胁和亲社会倾向

变量	亲社会倾向
	回归系数与显著性
感知机器人威胁	-0.077*** (0.007)
对机器人的一般态度	0.069*** (0.005)
机器人使用: 在家	0.024 (0.014)
机器人使用: 工作上	-0.026 (0.013)
机器人使用: 其他地方	-0.024 (0.027)
科学兴趣	控制
社会阶层	控制
教育	0.002*** (0.0003)
社区	控制
年龄	-0.000** (0.0002)
性别	0.052*** (0.006)
常量	2.614*** (0.043)
国家固定效应	控制
观测值	25769
调整 $R^2$	0.113
$F$	99.74***

注: ln GDP 和国家高共线性(VIF = 17.17), 因此在 HDFE 线性回归分析中删除 ln GDP, 使用 Esttab 比较剔除 ln GDP 前后两个模型的结果发现, 所有预测变量对因变量的回归系数和显著性不受影响, 因此在表中呈现剔除 ln GDP 后的结果。本分析还控制了个体的科学兴趣、社会阶层和社区(处理为虚拟变量, 因此受限于篇幅, 未在表中报告回归系数)。

\*\* $p < 0.01$ , \*\*\* $p < 0.001$ , 括号里为标准误。

考研究 1 的相关分析结果取中等效应量  $f = 0.30$ , 显著性水平  $\alpha = 0.05$ , 要达到 95% 的统计检验力至少需要 134 名被试。考虑到可能有被试中途退出或未通过注意检查, 在 Credamo 平台上共计招募了 150 名被试。其中 2 名被试没有通过注意检查, 最终被试量为 148 人。被试平均年龄为 30.11 岁 ( $SD = 5.71$ ), 其中女性 84 名(占 56.8%)。

## 4.2 材料

感知机器人威胁。采用 Yogeewaran 等人(2016)编制的感知机器人威胁量表。该量表有两个维度: 感知机器人现实威胁和感知机器人认同威胁, 每个维度包含 5 个条目。感知机器人现实威胁的条目如下: (1)“我们日常生活中机器人使用的增加正在导致人类失业”; (2)“机器人无法代替人们的工作”; (3)“从长远来看, 机器人对人类的安全和福祉构成直接威胁”; (4)“机器人技术的发展会威胁人类的就业和机会”; (5)“机器人在日常生活中的日益普及对人类安全构成了威胁”。其中条目 2 为反向计分。感知机器人认同威胁的条目如下: (1)机器人在日常生活中的广泛应用使我感到困扰, 因为它模糊了人类和机器之间的界限; (2)看起来有生命的机器人是

令人不安的, 因为它们与人类几乎无法区分; (3)技术的最新进步对人类的本质提出了挑战; (4)机器人领域的技术进步正威胁着人类的独特性; (5)机器人开始模糊人类与机器之间的界限。量表为 Likert7 点量表, 1 为非常不同意, 7 为非常同意。本研究中感知机器人现实威胁维度的 Cronbach's  $\alpha$  系数为 0.87, 感知机器人认同威胁维度的 Cronbach's  $\alpha$  系数为 0.93, 全部项目的 Cronbach's  $\alpha$  系数为 0.95。

亲社会倾向。对亲社会倾向的测量采用 Osgood 和 Muraven (2015)的亲社会倾向问卷。问卷包含 6 个条目: (1)帮助需要帮助的人是很重要的; (2)我想要帮助别人; (3)他人的幸福与福祉很重要; (4)让他人感到快乐是很重要的; (5)让所有人都感到快乐是很重要的; (6)他人的需求是重要的。问卷使用 Likert7 点计分, 1 为非常不同意, 7 为非常同意。得分越高表明亲社会倾向越强。本研究中该量表的 Cronbach's  $\alpha$  系数为 0.79。

集体焦虑。对集体焦虑的测量采用 Wohl 等人(2010)的集体焦虑量表。该量表包含 5 个条目: (1)我对人类的生存充满信心; (2)我对人类的未来有信心; (3)我对人类的未来感到担忧; (4)我认为人类将永远繁荣下去; (5)我担心人类未来的活力岌岌可危。其中条目 1、2、4 为反向计分。量表使用 Likert7 点计分, 1 为完全不同意, 7 为完全同意。得分越高表明集体焦虑越强。在本研究中, 全部条目的 Cronbach's  $\alpha$  系数为 0.92。

为了进一步增强研究结果的稳健性, 研究还加入了以下控制变量: (1)对机器人的熟悉程度(Leo & Huh, 2020): 你对机器人等人工智能的熟悉程度是? (1 = 完全不熟悉, 7 = 非常熟悉); (2)你对机器人等人工智能的喜爱程度是? (1 = 完全不喜欢, 7 = 非常喜欢); (3)主观社会阶层(Adler et al., 2000)。以及性别、年龄两项人口统计学变量。

## 4.3 结果

相关分析结果表明, 亲社会倾向与感知机器人威胁及其两个维度均呈显著负相关( $r_{威胁} = -0.41, p < 0.001$ ;  $r_{现实} = -0.40, p < 0.001$ ;  $r_{认同} = -0.40, p < 0.001$ ), 亲社会倾向与集体焦虑呈显著负相关( $r = -0.46, p < 0.001$ )。集体焦虑与感知机器人威胁及其两个维度均呈显著正相关( $r_{威胁} = 0.75, p < 0.001$ ;  $r_{现实} = 0.73, p < 0.001$ ;  $r_{认同} = 0.73, p < 0.001$ )。将机器人熟悉度、喜爱度、主观社会阶层和人口统计学变量作为控制变量进行偏相关分析, 结果发现各项相关仍然显著, 亲社会倾向与感知机器人威胁及其

两个维度均呈显著负相关( $r_{威胁} = -0.23, p = 0.005$ ;  $r_{现实} = -0.22, p = 0.008$ ;  $r_{认同} = -0.23, p = 0.006$ ), 亲社会倾向与集体焦虑呈显著负相关( $r = -0.26, p = 0.002$ )。集体焦虑与感知机器人威胁及其两个维度均呈显著正相关( $r_{威胁} = 0.60, p < 0.001$ ;  $r_{现实} = 0.59, p < 0.001$ ;  $r_{认同} = 0.57, p < 0.001$ )。

为了验证集体焦虑是否在感知机器人威胁对亲社会倾向的影响中起中介作用, 使用 SPSS 的 PROCESS 程序进行偏差校正的 Bootstrap 检验 (Hayes et al., 2013), 选择模型 4, 反复抽样 5000 次。在 95% 的置信区间下, 将感知机器人威胁作为自变量, 将亲社会倾向作为因变量, 将集体焦虑作为中介变量, 做中介效应分析。中介效应检验结果显示, 95% 的 Bootstrap 置信区间不包含 0, 因此集体焦虑中介了感知机器人威胁对亲社会倾向的影响(间接效应 =  $-0.12, SE = 0.05, 95\% CI [-0.218, -0.024]$ ), 在控制了集体焦虑后, 感知机器人威胁对亲社会倾向的直接影响不再显著(直接效应 =  $-0.06, SE = 0.05, 95\% CI [-0.161, 0.034]$ ), 说明集体焦虑起完全中介作用。

#### 4.4 讨论

研究 2 以问卷调查的方式进一步证明了感知机器人威胁与亲社会倾向的负向关系。此外, 研究 2 还支持了集体焦虑在感知机器人威胁与亲社会倾向的关系中的中介作用。但研究 2 也存在局限性, 问卷调查的结果可能受到共同因素干扰而无法确定因果关系。采用验证性因素分析考察共同方法偏差, 结果发现控制模型的  $\chi^2 = 334.5, df = 185$ , 与基线模型的  $\chi^2 = 362.5, df = 186$ , 差值为 28.000 ( $df = 1$ ), 达到显著水平( $p < 0.05$ ), 说明存在一定的共同方法偏差问题。因此, 研究 3 将以实验研究的方法检验感知机器人威胁对亲社会倾向的影响及集体焦虑的中介作用。

### 5 研究 3: 集体焦虑的中介作用(实验)

研究 3 用实验的方式操纵被试感知机器人威胁的水平, 一方面验证感知机器人威胁影响亲社会倾向的因果效应, 另一方面也再次验证集体焦虑的中介作用。

#### 5.1 被试

本研究通过软件 G\*Power 3.1 来确定所需样本量, 取中等效应量  $f = 0.2$ , 显著性水平  $\alpha = 0.05$ , 单因素两水平被试间设计需要共 266 名被试才能达到 90% 的统计检验力。在 Credamo 平台实时剔除没有

通过注意检查的被试数据并滚动采集, 剩余有效数据共 270 人。被试平均年龄为 29.01 岁 ( $SD = 7.07$ ), 女性 166 名 (占 61.5%)。所有被试在实验开始之前均仔细阅读了实验说明并知情同意, 有效数据被试在实验结束后获得一定实验报酬。

#### 5.2 程序

本研究为单因素两水平被试间设计, 被试被随机分配到高感知机器人威胁组或低感知机器人威胁组, 在最终有效被试中, 高感知机器人威胁组和低感知机器人威胁组分别 135 人。通过让两组被试观看不同的机器人视频<sup>1</sup>操纵其感知机器人威胁的高低水平 (Yogeeswaran et al., 2016; Zlotowski et al., 2017): 高感知机器人威胁组被试观看的视频中, 机器人在现实世界中的表现十分优异, 其行走、开门、搬运箱子等运动能力以及与人类传递物品、对话等交互能力均很强, 能够很好地完成各项任务; 而低感知机器人威胁组被试观看的视频则相反, 视频中机器人在现实世界中的表现非常糟糕, 能力低下, 无法很好地完成各项任务。被试观看完视频后填写了感知机器人威胁量表(感知机器人威胁的测量同研究 2, 在本实验中, 现实威胁维度的 Cronbach's  $\alpha$  系数为 0.84, 认同威胁维度的 Cronbach's  $\alpha$  系数为 0.91, 全部项目的 Cronbach's  $\alpha$  系数为 0.93)作为操纵检查。

由于机器人视频尤其是低威胁组机器人失败的视频可能会让被试感到比较有趣, 从而对实验结果产生一定影响, 因此被试接下来填写了感知视频乐趣量表, 该量表改编自 Agarwat 和 Karahanna (2000) 的研究, 共包含以下 4 个条目: (1) 看这个视频让我充满乐趣; (2) 看这个视频给我提供了很多快乐; (3) 我很享受这个视频; (4) 看这个视频让我感到无聊。其中条目 4 为反向计分。量表使用 Likert 7 点计分, 1 为非常不同意, 7 为非常同意。得分越高表明感知视频乐趣越强。在本研究中, 全部条目的 Cronbach's  $\alpha$  系数为 0.89。

然后进行集体焦虑测量(测量工具同研究 2, 在本实验中, 全部条目的 Cronbach's  $\alpha$  系数为 0.92)和亲社会倾向测量(测量工具同研究 2, 本实验中该量表的 Cronbach's  $\alpha$  系数为 0.89)。测量条目中穿插有注意检查题目, 最后填写人口统计学信息。

<sup>1</sup> 高威胁组视频链接: [https://www.bilibili.com/video/BV1p8411f7Rm/?spm\\_id\\_from=333.999.0.0&vd\\_source=7a98762f30b5656da1f201f5059a10b5](https://www.bilibili.com/video/BV1p8411f7Rm/?spm_id_from=333.999.0.0&vd_source=7a98762f30b5656da1f201f5059a10b5)

低威胁组视频链接: [https://www.bilibili.com/video/BV1K04y1M7hi/?spm\\_id\\_from=333.337.search-card.all.click](https://www.bilibili.com/video/BV1K04y1M7hi/?spm_id_from=333.337.search-card.all.click)

### 5.3 结果

单因素方差分析结果显示,高威胁组的感知机器人威胁( $M = 4.09, SD = 1.31, 95\% CI [3.86, 4.31]$ )显著高于低威胁组( $M = 3.51, SD = 1.32, 95\% CI [3.28, 3.73]$ ),  $F(1, 268) = 13.16, p < 0.001, \eta_p^2 = 0.47, 90\% CI [0.014, 0.093]$ , 表明操纵有效。

单因素方差分析结果显示,高威胁组的集体焦虑( $M = 3.15, SD = 1.32, 95\% CI [2.93, 3.37]$ )显著高于低威胁组( $M = 2.84, SD = 1.26, 95\% CI [2.62, 3.06]$ ),  $F(1, 268) = 4.01, p = 0.046, \eta_p^2 = 0.02, 90\% CI [0.000, 0.055]$ 。

单因素方差分析结果显示,高威胁组的亲社会倾向( $M = 5.20, SD = 0.96, 95\% CI [5.04, 5.37]$ )显著低于低威胁组( $M = 5.44, SD = 0.73, 95\% CI [5.32, 5.57]$ ),  $F(1, 268) = 5.21, p = 0.023, \eta_p^2 = 0.02, 90\% CI [0.001, 0.054]$ 。相关分析结果表明,感知机器人威胁程度(低威胁组 = 0, 高威胁组 = 1)与集体焦虑呈显著正相关( $r = 0.12, p = 0.046$ ),与亲社会倾向呈显著负相关( $r = -0.14, p = 0.023$ ),集体焦虑与亲社会倾向呈显著负相关( $r_{控制} = -0.39, p < 0.001$ )。将感知视频乐趣作为控制变量进行协方差分析,结果表明高威胁组的亲社会倾向仍然低于低威胁组,差异呈边缘显著,  $F(1, 267) = 3.58, p = 0.059, \eta_p^2 = 0.01$ 。

为了验证集体焦虑是否在感知机器人威胁对亲社会倾向的影响中起中介作用,使用 SPSS 的 PROCESS 程序进行偏差校正的 Bootstrap 检验(Hayes et al., 2013),选择模型 4,反复抽样 5000 次。在 95%的置信区间下,将感知机器人威胁作为自变量(低威胁组 = 0, 高威胁组 = 1),将亲社会倾向作为因变量,将集体焦虑作为中介变量,做中介效应分析。中介效应检验结果显示,95%的置信区间不包含 0,因此集体焦虑中介了感知机器人威胁对亲社会倾向的影响(间接效应 =  $-0.08, SE = 0.04, 95\% CI [-0.178, -0.004]$ ),在控制了集体焦虑后,感知机器人威胁对亲社会倾向的直接影响不再显著(直接效应 =  $-0.16, SE = 0.10, 95\% CI [-0.348, 0.033]$ ),说明集体焦虑起完全中介作用。以上结果表明,感知机器人威胁会降低亲社会倾向,这是由于感知机器人威胁会增强人们的集体焦虑,进而使人更不倾向于做出亲社会行为。

### 5.4 讨论

研究 3 通过实验研究的方式支持了感知机器人威胁影响亲社会倾向的因果效应,也再次支持了集体焦虑的中介作用。但感知机器人威胁对亲社会倾

向的影响也可能受到其他可能中介的影响,如泛人类主义(Jackson et al., 2020)、工作不安全感/事业焦虑(Yam et al., 2023)和负面情绪(Yam et al., 2023)。因此,研究 4 将检验其他可能中介的影响。

## 6 研究 4: 排除其它中介机制

为了进一步增强研究 3 实验结果的鲁棒性,研究 4 更换了自变量的操纵方式和因变量的测量方式,并且加入了其它可能中介变量的测量,以期再次验证集体焦虑的中介作用。研究 4 测量了泛人类主义、工作不安全感/事业焦虑和负面情绪(Yam et al., 2023)。泛人类主义是指人们感知到的与所有其他人类群体的亲近关系,无论种族、宗教或国籍,以往研究发现,机器人凸显通过泛人类主义减少群体间偏见(Jackson et al., 2020)。工作不安全感/失业焦虑指员工对其工作和职业未来的稳定性和安全性感到担忧或不安,以往研究发现,工作中接触机器人通过工作不安全感增加员工的工作倦怠与职场不文明行为(Yam et al., 2023)。

### 6.1 被试

本研究通过软件 G\*Power 3.1 来确定所需样本量,取中等效应量  $f = 0.2$ ,显著性水平  $\alpha = 0.05$ ,单因素两水平被试间设计需要共 266 名被试才能达到 90%的统计检验力。在 Credamo 平台实时剔除没有通过注意检查的被试数据并滚动采集,剩余有效数据共 270 人。被试平均年龄为 29.99 岁( $SD = 7.48$ ),女性 197 名(占 73.0%)。所有被试在实验开始之前均仔细阅读了实验说明并知情同意,有效数据被试在实验结束后获得一定实验报酬。

### 6.2 程序

本研究为单因素两水平被试间设计,被试被随机分配到机器人威胁组或对照组,在最终有效被试中,威胁组和对照组分别 135 人。为了对感知机器人威胁程度进行操纵,被试首先被要求根据一段指导语进行写作(许丽颖等, 2024),其中威胁组的指导语如下:

“随着人工智能的飞速发展,机器人将会给人类带来许多潜在威胁(如失业、不安全、挑战人类本质等)。请想象并描述你现在或将来可能面临的机器人威胁,以及被威胁之后你将会面临的情况(不少于 100 字)。”

对照组的指导语如下:

“随着人工智能的飞速发展,机器人将会协助人类进行许多社会活动(如送餐服务、物流运输、

教育辅助等)。请想象并描述一个未来机器人协助人类的场景(不少于100字)。”

被试在完成写作任务后填写了感知机器人威胁量表(感知机器人威胁的测量同研究2, 在本实验中, 现实威胁维度的 Cronbach's  $\alpha$  系数为 0.87, 认同威胁维度的 Cronbach's  $\alpha$  系数为 0.90, 全部项目的 Cronbach's  $\alpha$  系数为 0.93)作为操纵检查。

然后进行可能中介变量的测量。集体焦虑的测量同研究2, 在本实验中, 全部条目的 Cronbach's  $\alpha$  系数为 0.90。对泛人类主义的测量采用 McFarland 等(2012)编制的泛人类主义量表: 呈现给被试 5 组图形, 每组图形都有两个圆圈, 但圆圈的重叠程度各不相同; 其中一个圆圈代表“我”, 另一个圆圈则代表“人类”; 被试需要选择最能代表他们如何看待他们自身与整个人类群体之间关系的图形; 例如, 选择两个完全独立的圆圈表示被试认为自身与作为一个整体的人类隔绝, 而选择两个完全重叠的圆圈则表示被试与所有其他人类的极度亲密。对工作不安全感/事业焦虑的测量采用 Yam 等人(2023)的量表, 该量表包括 3 个条目: (1)我觉得我很快会失去我的工作; (2)我对我的工作前景感到不安; (3)我觉得我可能在不久的将来失去工作。量表使用 Likert 7 点计分, 1 为完全不同意, 7 为完全同意。得分越高表明工作不安全感/事业焦虑越强。在本研究中, 全部条目的 Cronbach's  $\alpha$  系数为 0.90。对负面情感的测量包括 6 个条目: (1)害怕; (2)沮丧; (3)焦虑; (4)恐惧; (5)忧虑; (6)生气。量表使用 Likert 7 点计分, 1 为完全没有, 7 为非常强烈, 得分越高表明负面情感越强。在本研究中, 全部条目的 Cronbach's  $\alpha$  系数为 0.92。

接下来测量被试的亲社会倾向, 本研究的亲社会倾向测量采用 Touré-Tillery 和 Light (2018)的场景测量量表, 被试需要阅读 8 个简短场景, 并表明其有多大可能参与描述的行为(1 = 绝对不会, 7 = 绝对会), 例如: “你正在赴约路上, 快迟到了, 这时有人走过来问你去一个你知道的地方的路。你会停下来给这个人指路吗? ”。得分越高表明亲社会倾向越强。在本研究中, 全部条目的 Cronbach's  $\alpha$  系数为 0.79。中介变量和因变量的测量条目中穿插有注意检查题目, 最后填写人口统计学信息。

### 6.3 结果

单因素方差分析结果显示, 威胁组的感知机器人威胁( $M = 4.83, SD = 0.97, 95\% CI [4.65, 5.01]$ )显著高于对照组( $M = 3.36, SD = 1.14, 95\% CI [3.18,$

$3.54]$ ),  $F(1, 268) = 130.51, p < 0.001, \eta_p^2 = 0.33, 90\% CI [0.254, 0.394]$ , 表明操纵有效。

单因素方差分析结果显示, 威胁组的亲社会倾向( $M = 4.86, SD = 0.96, 95\% CI [4.71, 5.01]$ )显著低于对照组( $M = 5.11, SD = 0.82, 95\% CI [4.95, 5.26]$ ),  $F(1, 268) = 5.25, p = 0.023, \eta_p^2 = 0.02, 90\% CI [0.001, 0.054]$ 。相关分析结果表明, 感知机器人威胁程度(对照组 = 1, 威胁组 = 2)与集体焦虑呈显著正相关( $r = 0.40, p < 0.001$ ), 与亲社会倾向呈显著负相关( $r = -0.14, p = 0.023$ ), 集体焦虑与亲社会倾向呈显著负相关( $r_{控制} = -0.39, p < 0.001$ )。

为了验证集体焦虑是否在感知机器人威胁对亲社会倾向的影响中起中介作用, 使用 SPSS 的 PROCESS 程序进行偏差校正的 Bootstrap 检验 (Hayes et al., 2013), 选择模型 4, 反复抽样 5000 次。在 95% 的置信区间下, 将感知机器人威胁作为自变量(对照组 = 1, 威胁组 = 2), 将亲社会倾向作为因变量, 将集体焦虑作为中介变量, 做中介效应分析。中介效应检验结果显示, 95% 的置信区间不包含 0, 因此集体焦虑中介了感知机器人威胁对亲社会倾向的影响(间接效应 =  $-0.29, SE = 0.06, 95\% CI [-0.427, -0.190]$ ), 在控制了集体焦虑后, 感知机器人威胁对亲社会倾向的直接影响不再显著(直接效应 =  $-0.04, SE = 0.11, 95\% CI [-0.175, 0.260]$ ), 说明集体焦虑起完全中介作用。以上结果再次支持了集体焦虑的中介作用, 即感知机器人威胁降低亲社会倾向是由于增强了人们的集体焦虑。为了排除其它可能中介变量, 用同上方式对泛人类主义、工作不安全感/事业焦虑、负面情感做中介效应分析。泛人类主义的中介效应检验结果显示, 95% 的置信区间包含 0, 因此泛人类主义无法中介感知机器人威胁对亲社会倾向的影响(间接效应 =  $-0.05, SE = 0.03, 95\% CI [-0.132, 0.004]$ )。工作不安全感的中介效应检验结果显示, 95% 的置信区间不包含 0, 因此工作不安全感中介了感知机器人威胁对亲社会倾向的影响(间接效应 =  $-0.13, SE = 0.05, 95\% CI [-0.235, -0.052]$ ), 在控制了工作不安全感后, 感知机器人威胁对亲社会倾向的直接影响不再显著(直接效应 =  $-0.12, SE = 0.12, 95\% CI [-0.344, 0.111]$ ), 说明工作不安全感起完全中介作用。负面情感的中介效应检验结果显示, 95% 的置信区间包含 0, 因此负面情感无法中介感知机器人威胁对亲社会倾向的影响(间接效应 =  $-0.07, SE = 0.05, 95\% CI [-0.167, 0.019]$ )。

由于工作不安全感中介效应显著，但其间接效应(-0.13)小于集体焦虑(-0.29)，因此其中介的程度比集体焦虑更小。为了进一步比较工作不安全感与集体焦虑在感知机器人威胁和亲社会倾向之间的中介效应，首先，采用回归分析方式考察二者的区别。以亲社会倾向作为因变量，感知机器人威胁(第一层)、工作不安全感与集体焦虑(第二层)作为自变量进行分层线性回归，结果发现，集体焦虑对亲社会倾向的预测作用显著(标准化  $\beta = -0.45, t = -5.80, p < 0.001$ )，而工作不安全感对亲社会倾向的预测作用不显著(标准化  $\beta = 0.08, t = 1.02, p = 0.307$ )。以亲社会倾向为因变量，感知机器人威胁为自变量，集体焦虑和工作不安全感为中介进行路径分析发现，模型中集体焦虑对亲社会倾向的路径系数为-0.325 ( $SE = 0.044, CR = -7.421, p < 0.001$ )，工作不安全感影响亲社会倾向的路径系数为 0.049 ( $SE = 0.038, CR = 1.310, p = 0.190$ )。在同一模型中对比可见，工作不安全感对亲社会倾向的影响比集体焦虑对亲社会倾向的影响更弱。

### 6.4 讨论

研究 4 进一步支持了感知机器人威胁影响亲社会倾向的因果效应，也再次支持了集体焦虑的中介作用。此外，研究 4 排除了泛人类主义和负面情感在感知机器人威胁对亲社会倾向的影响中的中介作用。虽然以往研究发现机器人凸显可以促进人们的泛人类主义(Jackson et al., 2020)，但这可能建立在威胁不突出的情况下。然而，当机器人所带来的威胁变得明显时，这种效应可能不存在。此外，研究 4 发现工作不安全感也能中介感知机器人威胁对亲社会倾向的影响，这与以往研究是一致的(Yam et al., 2023)。

## 7 研究 5: 内外群体的调节作用

研究 5 通过对内外群体的操纵探究其在感知机

器人威胁影响亲社会倾向中的调节作用。

### 7.1 被试

采用 G\*Power 3.1 软件(Faul et al., 2007)计算本实验所需样本量，对于本实验适用的双因素方差分析，取中等效应量  $f = 0.2$ ，显著性水平  $\alpha = 0.05$ ，组数为 4，要达到 90% 的统计检验力至少需要 359 名被试。通过 Credamo 平台招募被试，随机分配至内群体高威胁组、内群体低威胁组、外群体高威胁组和外群体低威胁组，实时剔除没有通过注意检查的被试数据并滚动采集，剩余有效数据共 400 名(女性 264 名)被试年龄在 19~60 岁( $M = 30.95, SD = 6.99$ )之间，其中每组均为 100 人。所有被试在实验开始之前均仔细阅读了实验说明并知情同意，有效数据被试在实验结束后获得一定实验报酬。

### 7.2 程序

实验 5 为 2 (感知机器人威胁: 高 vs. 低)  $\times$  2 (内外群体: 内 vs. 外) 被试间实验设计，所有被试被随机分配到 4 个组的其中一组。

首先对被试的感知机器人威胁水平进行操纵，参考相关研究(Jackson et al., 2020; 许丽颖等, 2024)的操纵材料，制作了两张不同的新闻网页图片(如图 1)。在高感知机器人威胁组，被试阅读了一篇名为“机器人：取代人类劳动力？”的科技新闻，新闻描述了机器人将会抢走人类的工作，对人类构成真正的威胁。在低感知机器人威胁组，被试阅读了一篇名为“机器人：只是一时流行？”的科技新闻，文章描述了机器人抢占人类工作的可能性很小，不会对人类构成真正的威胁。两个新闻网页的布局、新闻的长度和格式都基本相同。被试看完新闻材料后填写了机器人威胁量表作为操纵检查(测量同研究 2，在本研究中，感知机器人现实威胁维度的 Cronbach's  $\alpha$  系数为 0.94，感知机器人认同威胁维度的 Cronbach's  $\alpha$  系数为 0.94，全部项目的 Cronbach's  $\alpha$  系数为 0.96)。



图 1 感知机器人威胁的操纵材料

接下来测量被试对于内群体或外群体成员的亲社会倾向。内外群体组的被试都阅读了以下对于内外群体的概念解释:“在社会心理学中,内群体是指一个人经常参与的或在其间生活、或在其间工作、或在其间进行其他活动的群体。在群体中的成员会感到自己与群体的关系十分密切,并对群体有强烈的归属感。与其相对应的是外群体,外群体泛指人们没有参与也没有归属感的群体。”(Brewer, 1999) 内群体组的被试被要求写下一个他们认为属于其内群体的人的名字,而外群体组的被试被要求写下一个他们认为属于其外群体的人的名字(均为首字母缩写即可)。然后,请被试将写下的人名代入对于亲社会倾向的测量(测量工具同研究2,本实验中该量表的Cronbach's  $\alpha$ 系数为0.97),例如“帮助需要帮助的他/她(你在上题中所填的人)是很重要的。”并且,为了检验内外群体的操纵是否成功,被试还填写了内外群体测量,采用改编后的自我他人相容性量表(Aron et al., 1992)。被试被呈现7组图形,每组图形都有两个圆圈,但圆圈的重叠程度各不相同。其中一个圆圈代表“自己”,另一个圆圈则代表被试在上题中写下的“他/她”。被试被要求选择最能代表他们如何看待他们自己与所写之人之间关系的图形,两个圆圈重叠的程度越高,说明被试认为他们和所写之人的关系越近。

接着进行注意检查,以上量表中穿插了两道注意检查题目,如“本题请选择非常同意”。此外,还对操纵材料的内容进行了注意检查,题目如下:(1)根据研究,现在机器人占据的工作岗位比例(1为很大,2为很小);(2)根据这项研究作者的说法,机器人能力和智能的进步所需的时间将比预期(1为更长,2为更短)。最后,被试报告年龄和性别两项人口统计学信息。

### 7.3 结果

单因素方差分析结果显示,高威胁组的感知机器人威胁( $M = 4.93, SD = 1.33, 95\% CI [4.74, 5.11]$ )显著高于低威胁组( $M = 2.76, SD = 1.21, 95\% CI [2.59, 2.93]$ ),  $F(1, 398) = 292.47, p < 0.001, \eta_p^2 = 0.42$ ;内群体组的自我他人相容性( $M = 5.49, SD = 1.12, 95\% CI [5.33, 5.65]$ )显著高于外群体组( $M = 2.93, SD = 1.96, 95\% CI [2.65, 3.20]$ ),  $F(1, 398) = 257.77, p < 0.001, \eta_p^2 = 0.39$ ,表明操纵均有效。

以感知机器人威胁(低威胁组 = 0,高威胁组 = 1)和内外群体(内群体组 = 0,外群体组 = 1)作为自变量,以亲社会倾向作为因变量进行方差分析。

数据结果表明,内群体组的亲社会倾向评分( $M = 5.80, SD = 0.69, 95\% CI [5.70, 5.90]$ )显著高于外群体组( $M = 4.01, SD = 1.73, 95\% CI [3.76, 4.25]$ ),  $F(1, 396) = 188.16, p < 0.001, \eta_p^2 = 0.32$ ,低威胁组的亲社会倾向评分( $M = 5.02, SD = 1.51, 95\% CI [4.81, 5.23]$ )高于高威胁组( $M = 4.78, SD = 1.67, 95\% CI [4.55, 5.02]$ ),差异呈边缘显著,  $F(1, 396) = 3.31, p = 0.070, \eta_p^2 = 0.01$ ,感知机器人威胁和内外群体的交互作用显著,  $F(1, 396) = 3.99, p = 0.046, \eta_p^2 = 0.01$ 。简单效应分析发现,在外群体组,高威胁组的亲社会倾向评分( $M = 3.76, SD = 1.71, 95\% CI [3.50, 4.01]$ )显著低于低威胁组( $M = 4.26, SD = 1.72, 95\% CI [4.00, 4.51]$ ),  $F(1, 396) = 7.29, p = 0.007, \eta_p^2 = 0.02$ ;在内群体组,高低威胁组的亲社会倾向评分无显著差异,  $F(1, 396) = 0.02, p = 0.900, \eta_p^2 < 0.001$ (见图2)。

### 7.4 讨论

研究5进一步支持了感知机器人威胁对亲社会倾向的影响。此外,研究5还支持了内外群体的调节作用。对于外群体成员,机器人威胁显著降低了亲社会倾向,然而,机器人威胁并未提升对内群体成员的亲社会倾向。这可能是因为我们对内群体的操纵为让被试写下一个属于内群体的名字而非一个群体,所以被试可能写下了自己最亲近的人,从而无论是否受到机器人威胁,都具有很高的亲社会倾向。

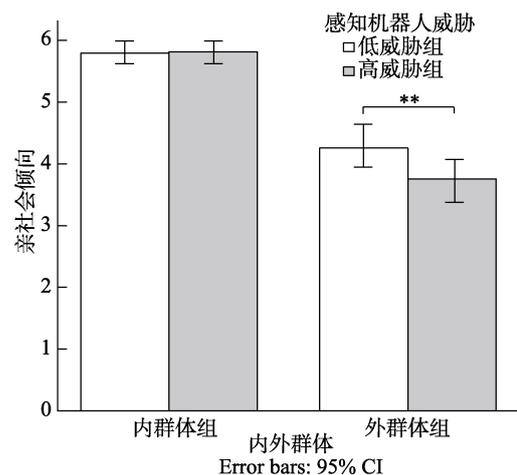


图2 不同感知机器人威胁组对内外群体成员的亲社会倾向评分

注: \*\* $p < 0.01$ .

## 8 研究6: 道德比较倾向的调节作用

研究6通过对向上和向下道德比较倾向的测量

探究道德比较倾向在感知机器人威胁影响亲社会倾向中的调节作用。

### 8.1 被试

本研究通过软件 G\*Power 3.1 来确定所需样本量,取中等效应量  $f = 0.2$ , 显著性水平  $\alpha = 0.05$ , 单因素两水平被试间设计需要共 266 名被试才能达到 90% 的统计检验力。考虑到可能会有被试无法完成实验或没有通过注意检查,在 Credamo 平台上共计招募了 297 名被试。其中 27 名被试没有通过注意检查,最终被试量为 270 人。被试平均年龄为 32.62 岁 ( $SD = 8.79$ ), 女性 155 名 (占 57.4%)。

### 8.2 程序

本研究为单因素两水平被试间设计,被试被随机分配到高威胁组和低威胁组,在最终有效被试中,高威胁组和低威胁组的被试均为 135 人。对于感知机器人威胁的操纵方式同研究 5。被试看完新闻材料后填写了机器人威胁量表作为操纵检查(测量同研究 2,在本研究中,感知机器人现实威胁维度的 Cronbach's  $\alpha$  系数为 0.95,感知机器人认同威胁维度的 Cronbach's  $\alpha$  系数为 0.95,全部项目的 Cronbach's  $\alpha$  系数为 0.97)。

接下来,依次测量了被试的亲社会倾向(测量工具同研究 2,本实验中该量表的 Cronbach's  $\alpha$  系数为 0.82)和道德比较倾向。对道德比较倾向的测量采用 Fleischmann 等人(2021)的道德比较倾向问卷,该问卷包括 6 个小故事,其中三个故事是测量向上道德比较倾向:(1)给奶奶惊喜;(2)帮助无家可归的年轻人;(3)在医院做志愿者。另外三个故事是测量向下道德比较倾向:(1)考试作弊;(2)欺骗妻子;(3)对朋友撒谎。对被试以随机顺序呈现 6 个小故事的标题,然后请被试选择他们对这些小故事感兴趣的程度。测量采用 Likert 7 点量表,1 为一点也不感兴趣,7 为非常感兴趣。将测量向下道德比较倾向的三个条目反向计分,问卷得分越高表明越倾向于向上道德比较,得分越低表明越倾向于向下道德比较。本研究中该问卷的 Cronbach's  $\alpha$  系数为 0.80。最后,被试进行了注意检查(同研究 3)并报告了人口统计学信息。

### 8.3 结果

单因素方差分析结果显示,高威胁组的感知机器人威胁 ( $M = 5.14, SD = 1.29, 95\% CI [4.92, 5.36]$ ) 显著高于低威胁组 ( $M = 2.68, SD = 1.26, 95\% CI [2.46, 2.89]$ ),  $F(1, 268) = 251.53, p < 0.001, \eta_p^2 = 0.48$ , 表明操纵有效。高威胁组的亲社会倾向 ( $M =$

$5.60, SD = 0.79, 95\% CI [5.47, 5.74]$ ) 显著低于低威胁组 ( $M = 5.79, SD = 0.64, 95\% CI [5.68, 5.90]$ ),  $F(1, 268) = 4.82, p = 0.029, \eta_p^2 = 0.02$ 。

以亲社会倾向为因变量考察感知机器人威胁(低威胁组 = -1, 高威胁组 = 1)与道德比较倾向的交互作用,结果表明感知机器人威胁和道德比较倾向对亲社会倾向存在边缘显著的交互作用 ( $b = 0.07, SE = 0.04, t = 1.89, p = 0.060$ ), 高威胁组的亲社会倾向低于低威胁组 ( $b = -0.12, SE = 0.04, t = -3.47, p = 0.001$ ), 向上道德比较倾向越强,亲社会倾向越强 ( $b = 0.44, SE = 0.04, t = 11.76, p < 0.001$ ), 模型的  $\Delta R^2 = 0.01, F(1, 266) = 3.56, p = 0.060$ 。交互作用图 3 所示,简单斜率分析结果表明,在向下道德比较倾向较强的条件下,感知机器人威胁对亲社会倾向的影响显著 ( $b = -0.19, SE = 0.05, t = -3.79, p < 0.001$ ); 而在向上道德比较倾向较强的条件下,感知机器人威胁对亲社会倾向的影响不显著 ( $b = -0.06, SE = 0.05, t = -1.12, p = 0.266$ )。

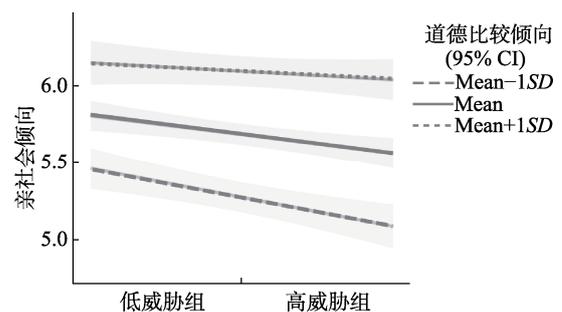


图 3 道德比较倾向的调节作用

### 8.4 讨论

研究 6 进一步支持了感知机器人威胁对亲社会倾向的负向影响。此外,研究 6 还部分发现了道德比较倾向的调节作用,机器人威胁显著降低了道德下行比较者的亲社会倾向,但并不会影响道德上行比较者的亲社会倾向。虽然研究 1~6 都发现了感知机器人威胁对亲社会倾向的影响,以及研究 4 验证了感知机器人威胁对自我报告的亲社会行为的影响,但可能存在自我报告的偏差。因此,研究 7 旨在进一步探讨感知机器人威胁对真实亲社会行为的影响。

## 9 研究 7: 感知机器人威胁对亲社会行为的影响(实验)

研究 7 以实地调查的方式探讨感知机器人威胁对亲社会行为的影响。

## 9.1 被试

本研究在一所知名大学中实地采集数据。研究共招募164名被试,其中3名被试未通过注意检查,最终样本量为161人。被试平均年龄为20.52岁( $SD = 1.68$ ),女性83名(占51.6%)。所有被试自愿参加实验并知情同意。

## 9.2 程序

本研究为单因素两水平被试间设计,被试被随机分配到机器人威胁组或对照组。为了对感知机器人威胁程度进行操纵,被试首先被要求根据一段指导语进行写作(许丽颖等,2024),研究范式同研究4。

被试在完成写作任务后填写了感知机器人威胁量表(感知机器人威胁的测量同研究2,在本实验中,现实威胁维度的Cronbach's  $\alpha$ 系数为0.65,认同威胁维度的Cronbach's  $\alpha$ 系数为0.75,全部项目的Cronbach's  $\alpha$ 系数为0.79)作为操纵检查。

然后,亲社会行为的测量参照Gaesser等(2020)的范式,要求被试给需要帮助的人写一封鼓励信,以信的字数作为亲社会行为的测量。指导语如下:

“在偏远贫困地区,许多孩子都面临着严峻的求学困难。由于各种原因,如交通不便、学校资源匮乏、家庭贫困等,他们可能很难获得良好的教育机会,甚至无法完成基本的学业。除了求学方面的困境,他们还可能面临着心理上的压力和挣扎,感到孤独与焦虑、自卑与自我怀疑。

在这样的困境中,这些孩子更加需要外界的关怀和支持。因此,我们邀请你给偏远贫困地区的孩子写一封鼓励信,这封信将给他们带来鼓励和希望,也将展现你的善心与温暖。你可以分享一些鼓励的话语,或者讲述你自己或身边人的学习经历和成长故事,希望能够激励这些孩子坚持努力、追求梦想。你的鼓励信将是给他们带来鼓励和希望的重要一步,感谢你的参与和支持!(字数不限)”

最后,收集了被试的性别和年龄两项人口统计学变量。

## 9.3 结果

单因素方差分析结果显示,高威胁组的感知机器人威胁( $M = 4.20, SD = 0.89, 95\% CI [4.02, 4.38]$ )显著高于低威胁组( $M = 3.88, SD = 0.82, 95\% CI [3.68, 4.07]$ ),  $F(1, 159) = 5.65, p = 0.019, \eta_p^2 = 0.34$ ,表明操纵有效。

单因素方差分析结果显示,高威胁组的鼓励信字数( $M = 71.74, SD = 122.52, 95\% CI [39.32,$

$104.15]$ )显著低于低威胁组( $M = 121.09, SD = 182.67, 95\% CI [85.97, 156.24]$ ),  $F(1, 159) = 4.16, p = 0.043, \eta_p^2 = 0.25$ 。即高威胁组的亲社会行为显著低于低威胁组。高威胁组的拟人化倾向( $M = 4.33, SD = 1.22, 95\% CI [4.08, 4.59]$ )与低威胁组( $M = 4.39, SD = 1.16, 95\% CI [4.12, 4.67]$ )没有显著差异,  $F(1, 159) = 0.10, p = 0.756, \eta_p^2 = 0.00$ 。高威胁组对机器人的能动性感知( $M = 4.62, SD = 1.05, 95\% CI [4.40, 4.85]$ )与低威胁组( $M = 4.39, SD = 1.09, 95\% CI [4.15, 4.64]$ )没有显著差异,  $F(1, 159) = 1.90, p = 0.170, \eta_p^2 = 0.01$ 。高威胁组对机器人的体验性感知( $M = 2.13, SD = 1.21, 95\% CI [1.88, 2.39]$ )与低威胁组( $M = 2.20, SD = 1.16, 95\% CI [1.93, 2.48]$ )也没有显著差异,  $F(1, 159) = 0.13, p = 0.715, \eta_p^2 = 0.00$ 。将拟人化倾向、感知能动性和感知体验性作为控制变量进行协方差分析,结果表明高威胁组的亲社会行为仍然低于低威胁组,差异呈边缘显著,  $F(1, 156) = 3.46, p = 0.065, \eta_p^2 = 0.02$ ,说明感知机器人威胁减少亲社会行为。

## 9.4 讨论

研究7以实地调查的方式进一步支持了感知机器人威胁不仅降低自我报告的亲社会倾向,也减少真实的亲社会行为。

## 10 总讨论

通过8个递进研究,我们发现了感知机器人威胁会降低人们的亲社会倾向,这一结论无论是在大规模调查数据库与国家层面的对应(研究1a, 1b)上,还是相关(研究2)及实验(研究3-7)中均得到了检验。并且,这种效应受到内外群体的调节,即虽然感知机器人威胁会降低人们的亲社会倾向,但这主要发生在外群体成员上,而若是面对内群体成员则此效应会减弱(研究5)。同时,道德比较的方向也会调节此效应,对于倾向于进行下行道德比较的人来说,感知机器人威胁会降低亲社会倾向;但对于倾向于上行道德比较者来说,这种影响则不显著(研究6)。此外,我们还发现感知机器人威胁也会影响真实的亲社会行为(研究7)。

本文的主要研究结果回应了机器人进入人类社会,尤其是对人们造成威胁会如何影响人际关系的争议。首先,我们发现感知机器人威胁会降低人们的亲社会倾向,这为以往研究所发现的感知机器人威胁会产生消极人际后果提供了新证据,也为丰富机器人群际威胁理论的人际后效研究提供了新

视角。机器人威胁会导致人机关系的恶化几乎已成共识,学者们从负面态度和使用倾向(Huang et al., 2021)、信任(Lin et al., 2022)、对后续研发投入的支持(Yogeeswaran et al., 2016)等不同角度均得到了趋于一致的结论。然而,关于感知机器人威胁对人际关系的影响究竟是好是坏仍未达成一致。持乐观态度的学者认为机器人的凸显会强化人类共同身份,从而带来减少偏见的积极人际后果(Jackson et al., 2020)。但似乎更多研究证据发现的则是消极后果,如在同一文章中作者也提到从 37 个国家收集的数据表明,在过去 42 年里,自动化发展最快的国家对外群体的外显偏见也有所增加(Jackson et al., 2020);也有研究发现人们对自动化兴起的担忧会导致对移民的负面态度和歧视行为(Gamez-Djokic & Waytz, 2020),工作场所机器人的威胁也会增加人们对同事的不文明行为(Yam et al., 2023)和物化(许丽颖等, 2024)。本文聚焦于亲社会倾向这一人际关系中的关键变量,不仅为机器人人际威胁理论的人际后效研究提供了重要视角,也直接回应了感知机器人威胁对人际关系影响的理论和现实争议。

其次,本文发现集体焦虑是感知机器人威胁影响亲社会倾向的中介机制,这不仅为感知机器人威胁的消极人际后果提供了新的理论解释,也进一步拓宽了集体焦虑后效的研究领域。以往研究对感知机器人威胁的人际影响研究多从不安全感(如 Gamez-Djokic & Waytz, 2020; Im et al., 2019)、控制补偿(许丽颖等, 2024)等个体视角进行解释,而我们的研究则选取了集体焦虑这个群体变量,为解释感知机器人威胁为何会导致消极人际结果提供了群体视角的解释。此外,关于集体焦虑的研究大多局限于人类群体,且其产生的人际后果也具有正反两面性(Wohl et al., 2012)。本文则将集体焦虑的研究延展至人工智能领域,发现机器人威胁也会引发人们的集体焦虑,拓展了集体焦虑研究的边界。此外,本研究发现,这种集体焦虑并未提升亲社会倾向,而是会降低人们的亲社会倾向。一种可能的解释是,虽然机器人威胁引发了集体焦虑,但并未使人类团结在一起,而是触发了面对共同威胁时的“替罪羊”心态(Brewer, 1999)。另一种可能的解释是,当人们面对外部的共同威胁而产生集体焦虑时,通常会通过对抗外群体来保护内群体利益。但当外部威胁来自机器人时,人们并未将机器人视为同人类一样具有完全心智的行为主体(Gray et al., 2007),因此也无法完全将其视为外群体对抗,而仍在人类

中区分外群体并进行消极人际互动。最后,由于对感知机器人威胁可能的负面人际后果及其心理机制的发现,本研究在实践上也具有在人工智能时代下的前瞻性警示作用。

同时,本文还发现了内外群体和道德比较倾向在感知机器人威胁影响亲社会倾向中的调节作用。首先,内外群体的调节作用或许可以为解释以往研究的分歧提供新的思路。如先前研究之所以会发现机器人会降低群体间偏见,是因为人们将机器人视为外群体,而将所有人都看作内群体,因此机器人所造成的积极抑或消极人际结果实际上是由于人心之群体划分所致(Jackson et al., 2020)。如果按照此正面效应的逻辑推断,则感知机器人威胁也应该增加而非降低亲社会倾向,因为若将所有人看作一个内群体,我们应该会更多帮助内群体成员,但为何我们研究的结果却恰恰相反呢?这可能有两方面的原因,一方面先前研究所探索的变量为机器人凸显而非威胁(Jackson et al., 2020),凸显并不一定会造成诸如威胁这种负面感知,且在程度上也会有所差异,这可能是造成结果迥异的原因之一;另一方面感知机器人威胁会造成一种集体焦虑,这一过程抽象且解释水平高,它并不依赖具体情境而产生情绪状态,但人们在做亲社会倾向的判断时,则是一种较低解释水平且具体化情境的判断,当人们被问到帮助他人意愿的问题时,人们可能已经在一定程度上将“他人”视为了外群体(Peng et al., 1997),这导致人们虽然会在感受到集体焦虑时将整个人类看作整体,但在考虑帮助他人的问题时又很快将“他人”视为外群体。正如研究 5 中我们发现,一旦明晰内外群体的划分,感知机器人威胁降低亲社会倾向的效应就不会在内群体中出现。但我们并未观察到感知机器人威胁对内群体成员亲社会倾向的提升影响。一种可能的解释是我们对内群体的操纵为让被试写下一个属于内群体的名字而非一个群体,所以被试可能写下了自己最亲近的人,从而无论是否受到机器人威胁,都具有很高的亲社会倾向。而如果让被试写下一个所属的群体,可能就会得到不同的结果。因为针对最亲近的人,即使在没有受到机器人威胁的情况下,亲社会倾向已经较高;但是对于所属内群体而言,在没有威胁的情况下,这种内群体依恋与偏好并不是特别突出,而在受到威胁时,内群体意识才会更加凸显(Brewer, 1999, p. 431; Sherif, 1966; Sumner, 1906)。未来研究可以检验感知机器人威胁对内群体亲社会倾向的影响是

否在个体层面和群体层面存在差异。

其次,本研究也在不同理论水平上探索了机器人威胁导致亲社会行为倾向减少的边界。内外群体是一种横向群体比较,而道德比较则是一种典型的纵向自我比较。研究6发现,感知机器人威胁会降低亲社会倾向只对倾向于进行下行道德比较者有效,而对上行比较者无效。这一结果出现的原因是进行下行道德比较会产生道德许可,从而允许自己放纵而做出一些不道德行为(Merritt et al., 2010)。当然,除了这一解释路径以外,这一研究结果或许也与向下比较时所造成的动机水平下降有关。先前研究表明,与更高的标准进行比较(即向上比较)能够形成一种激励作用,从而提升人们继续努力的动机水平(Diel & Hofmann, 2019);而与更低的标准进行比较(即向下比较)则会让人们满足于已有的进展,从而降低人们进一步努力的动机水平,这一过程被称为下滑(coasting)(Fulford et al., 2010)。下滑虽然在一定程度上能够增强自尊,但却并不会激励人们继续做出积极主动的行为,反而还会减少人们做出包括亲社会行为在内的积极行为的动机(Diel et al., 2021)。因此,向下比较所造成的下滑也可能是导致感知机器人威胁降低亲社会倾向这一效应在下行道德比较者中尤为显著的可能原因。从实践角度而言,我们对于内外群体和道德比较倾向这两个调节变量的发现,对于在现实生活中如何预防或削弱感知机器人威胁会造成的负面人际后果具有借鉴意义,例如可以通过对人们内外群体划分和道德比较倾向的干预达到弱化消极后果的作用。

当然,尽管本文通过系列研究对感知机器人威胁影响亲社会倾向的效应及其内在机制和边界条件进行了有益探索,但仍存在一定的局限性,有待未来继续探究。首先,根据实验因果链方法的逻辑,必须进一步操纵中介变量(即集体威胁),以检验其对因变量(即亲社会倾向)的影响(Spencer et al., 2005; Yang et al., 2024)。未来的研究应探索合适的实验范式,以操纵集体威胁并验证其对亲社会倾向的影响。第二,本研究主要考察了机器人威胁对外群体的亲社会态度和行为,未来研究应更多考虑机器人威胁对内群体亲社会行为的影响,探究机器人威胁是否会对内群体产生建设性影响。第三,本文探讨了感知机器人威胁对亲社会倾向的影响,但对机器人威胁的两个维度没有进行深入剖析。感知机器人威胁分为现实威胁和认同威胁两个维度,虽然这两个维度总是高度相关,且都会导致对机器人的

负面态度(如 Zlotowski et al., 2017),但二者的影响可能存在一定差异,如认同威胁更能预测人们对机器人的负面态度(Huang et al., 2021)和职场物化(许丽颖 等, 2024)。因此,未来研究可以对感知机器人威胁的两个维度进行区分研究,探索现实威胁和认同威胁对亲社会倾向的影响是否存在差异。第四,本文对亲社会的研究虽然在国家层面上涉及到了具体亲社会行为,但在个体水平上仍主要探讨亲社会倾向。虽然亲社会倾向对于亲社会行为有显著的预测作用,但倾向终究不是行为,并且亲社会行为也分为助人、合作、慈善捐赠、志愿服务等诸多种类(Penner et al., 2005),因此后续研究可以通过实验室实验、现场实验等方式对于不同的亲社会行为进行探究,进一步验证感知机器人威胁与实际亲社会行为之间的关联。最后,本研究发现了机器人威胁通过集体焦虑负面地影响了人际互动,即降低了人们的亲社会倾向。而以往研究发现,集体焦虑也可能促进群体间的团结。未来研究可进一步探索机器人威胁的何种性质使其产生的集体威胁降低人们的亲社会性,以及是否存在可逆转机器人威胁对亲社会负面影响的因素(如威胁程度、是否损害个人利益、资源稀缺性等)。

## 11 结论

本研究主要结论如下:第一,感知机器人威胁会降低亲社会倾向;第二,此效应受到集体焦虑的中介,即感知机器人威胁会增加集体焦虑,从而降低亲社会倾向;第三,此效应受到内外群体的调节,即感知机器人威胁主要会降低对于外群体成员的亲社会倾向;第四,此效应受到道德比较倾向的调节,即感知机器人威胁主要会降低下行道德比较者的亲社会倾向。

## 参 考 文 献

- Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy white women. *Health Psychology, 19*(6), 586-592.
- Agarwat, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly, 24*(4), 665-694.
- Agrawal, A., Gans, J., & Goldfarb, A. (2019). *The economics of artificial intelligence: An agenda*. Chicago and London: The University of Chicago Press.
- Andrighetto, L., Vezzali, L., Bergamini, G., Nadi, C., & Giovannini, D. (2016). Inside the earthquake: Perceived

- disaster exposure and helping intentions among Italian and immigrant victims of the 2012 Italian earthquakes. *Group Processes and Intergroup Relations*, 19(6), 753–768.
- Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63(4), 596–612.
- Asimov, I. (1950). *I, Robot*. New York: Gnome Press.
- Bai, M., Zhang, H., Zhang, J., Jiang, Y., & Xu, J. (2025). Challenging or threatening? The double-edged sword effect of intelligent technology awareness on accountants' unethical decision-making. *Journal of Business Ethics*, 197, 159–175.
- Barnes, C. M., Dang, C. T., Leavitt, K., Guarana, C. L., & Uhlmann, E. L. (2018). Archival data in micro-organizational research: A toolkit for moving to a broader set of topics. *Journal of Management*, 44(4), 1453–1478.
- Batson, C. D., & Powell, A. A. (2003). Altruism and prosocial behavior. In T. Millon & M. J. Lerner (Eds.), *Handbook of psychology: Personality and social psychology* (Vol. 5, pp. 463–484). Hoboken, NJ: Wiley.
- Bavel, J. J. V., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., ... Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, 4(5), 460–471.
- Bobo, L. (1983). Whites' opposition to busing: Symbolic racism or realistic group conflict? *Journal of Personality and Social Psychology*, 45(6), 1196–1210.
- Bordot, F. (2022). Artificial intelligence, robots and unemployment: Evidence from OECD countries. *Journal of Innovation Economics and Management*, 37(1), 117–138.
- Bostrom, N. (2002). Existential risks: Analyzing human extinction scenarios and related hazards. *Journal of Evolution and Technology*, 9, 1–30.
- Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate?. *Journal of Social Issues*, 55(3), 429–444.
- Brewer, M. B. (2001). The many faces of social identity: Implications for political psychology. *Political Psychology*, 22(1), 115–125.
- Butz, D. A., & Yogeewaran, K. (2011). A new threat in the air: Macroeconomic threat increases prejudice against Asian Americans. *Journal of Experimental Social Psychology*, 47(1), 22–27.
- Carradore, M. (2022). People's attitudes towards the use of robots in the social services: A multilevel analysis using eurobarometer data. *International Journal of Social Robotics*, 14, 845–858.
- Cortland, C. I., Craig, M. A., Shapiro, J. R., Richeson, J. A., Neel, R., & Goldstein, N. J. (2017). Solidarity through shared disadvantage: Highlighting shared experiences of discrimination improves relations between stigmatized groups. *Journal of Personality and Social Psychology*, 113(4), 547–567.
- Diel, K., Grelle, S., & Hofmann, W. (2021). A motivational framework of social comparison. *Journal of Personality and Social Psychology*, 120(6), 1415–1430.
- Diel, K., & Hofmann, W. (2019). Inspired to perspire: The interplay of social comparison direction and standard extremity in the context of challenging exercising goals. *Social Cognition*, 37(3), 247–265
- Dovidio, J. F., ten Vergert, M., Stewart, T. L., Gaertner, S. L., Johnson, J. D., Esses, V. M., Riek, B. M., & Pearson, A. R. (2004). Perspective and prejudice: Antecedents and mediating mechanisms. *Personality and Social Psychology Bulletin*, 30(12), 1537–1549.
- Drury, J. (2018). The role of social identity processes in mass emergency behaviour: An integrative review. *European Review of Social Psychology*, 29(1), 38–81.
- 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.
- Fleischmann, A., Lammers, J., Diel, K., Hofmann, W., & Galinsky, A. D. (2021). More threatening and more diagnostic: How moral comparisons differ from social comparisons. *Journal of Personality and Social Psychology*, 121(5), 1057–1078.
- Fleming, D. A., Chong, A., & Bejarano, H. D. (2014). Trust and reciprocity in the aftermath of natural disasters. *The Journal of Development Studies*, 50(11), 1482–1493.
- Ford, M. R. (2009). *The lights in the tunnel: Automation, accelerating technology and the economy of the future*. Wayne, PA: Acculant Publishing.
- Frankenberg, E., Nobles, J., & Sumantri, C. (2012). Community destruction and traumatic stress in post-tsunami Indonesia. *Journal of Health and Social Behavior*, 53(4), 498–514.
- Frey, C. B., Berger, T., & Chen, C. (2018). Political machinery: Did robots swing the 2016 US presidential election? *Oxford Review of Economic Policy*, 34(3), 418–442.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation?. *Technological Forecasting and Social Change*, 114, 254–280.
- Fuesting, M. A., Diekmann, A. B., Boucher, K. L., Murphy, M. C., Manson, D. L., & Safer, B. L. (2019). Growing STEM: Perceived faculty mindset as an indicator of communal affordances in STEM. *Journal of Personality and Social Psychology*, 117(2), 260–281.
- Fulford, D., Johnson, S. L., Llabre, M. M., & Carver, C. S. (2010). Pushing and coasting in dynamic goal pursuit: Coasting is attenuated in bipolar disorder. *Psychological Science*, 21(7), 1021–1027.
- Gaesser, B., Shimura, Y., & Cikara, M. (2020). Episodic simulation reduces intergroup bias in prosocial intentions and behavior. *Journal of Personality and Social Psychology*, 118(4), 683–705.
- Gamez-Djokic, M., & Waytz, A. (2020). Concerns about automation and negative sentiment toward immigration. *Psychological Science*, 31(8), 987–1000.
- Gordils, J., Elliot, A. J., Toprakiran, S., & Jamieson, J. P. (2021). The effects of COVID-19 on perceived intergroup competition and negative intergroup outcomes. *The Journal of Social Psychology*, 161(4), 419–434.
- Goyal, A., & Aneja, R. (2020). Artificial intelligence and income inequality: Do technological changes and worker's position matter?. *Journal of Public Affairs*, 20(4), e2326.
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315(5812), 619.
- Gray, K., Yam, K. C., Zhen'An, A. E., Wilbanks, D., & Waytz, A. (2023). The psychology of robots and artificial intelligence. In D. Gilbert et al (Eds.), *Handbook of social psychology* (pp. 1–83). Situational Press.
- Harari, Y. N. (2017). Reboot for the AI revolution. *Nature*, 550(7676), 324–327.
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.
- Hirschberger, G., & Pyszczynski, T. (2011). Killing with a

- clean conscience: Existential angst and the paradox of morality. In M. Mikulincer & P. R. Shaver (Eds.), *Social psychology of morality: Exploring the causes of good and evil* (pp. 331–347). Washington, DC: American Psychological Association.
- Howard, A., & Borenstein, J. (2018). The ugly truth about ourselves and our robot creations: The problem of bias and social inequity. *Science and Engineering Ethics, 24*(5), 1521–1536.
- Huang, H. L., Cheng, L. K., Sun, P. C., & Chou, S. J. (2021). The effects of perceived identity threat and realistic threat on the negative attitudes and usage intentions toward hotel service robots: The moderating effect of the robot's anthropomorphism. *International Journal of Social Robotics, 13*, 1599–1611.
- Im, Z. J., Mayer, N., Palier, B., & Rovny, J. (2019). The “losers of automation”: A reservoir of votes for the radical right? *Research & Politics, 6*(1), 1–7.
- Jackson, J. C., Castelo, N., & Gray, K. (2020). Could a rising robot workforce make humans less prejudiced? *American Psychologist, 75*(7), 969–982.
- Jackson, J. W. (1993). Realistic group conflict theory: A review and evaluation of the theoretical and empirical literature. *Psychological Record, 43*(3), 395–413.
- Jetten, J., Mols, F., Healy, N., & Spears, R. (2017). “Fear of falling”: Economic instability enhances collective angst among societies' wealthy class. *Journal of Social Issues, 73*(1), 61–79.
- Jetten, J., Mols, F., & Steffens, N. K. (2021). Prosperous but fearful of falling: The wealth paradox, collective angst, and opposition to immigration. *Personality and Social Psychology Bulletin, 47*(5), 766–780.
- Jetten, J., & Wohl, M. J. A. (2012). The past as a determinant of the present: Historical continuity, collective angst, and opposition to immigration. *European Journal of Social Psychology, 42*(4), 442–450.
- Kahn, B. (2023). *Elon Musk and Apple cofounder Steve Wozniak among over 1,100 who sign open letter calling for 6-month ban on creating powerful A.I.* Retrieved August 17, 2024, from <https://fortune.com/2023/03/29/elon-musk-apple-steve-wozniak-over-1100-sign-open-letter-6-month-ban-creating-powerful-ai/>
- Kahn, D. T., Björklund, F., & Hirschberger, G. (2022). The intent and extent of collective threats: A data-driven conceptualization of collective threats and their relation to political preferences. *Journal of Experimental Psychology: General, 151*(5), 1178–1198.
- Khasawneh, O. Y. (2018). Technophobia: Examining its hidden factors and defining it. *Technology in Society, 54*, 93–100.
- Kung, F. Y., Chao, M. M., Yao, D. J., Adair, W. L., Fu, J. H., & Tasa, K. (2018). Bridging racial divides: Social constructionist (vs. essentialist) beliefs facilitate trust in intergroup contexts. *Journal of Experimental Social Psychology, 74*, 121–134.
- Lalonde, R. N. (2002). Testing the social identity-intergroup differentiation hypothesis: We're not American eh! *British Journal of Social Psychology, 41*(4), 611–630.
- Lasarov, W., & Hoffmann, S. (2020). Social moral licensing. *Journal of Business Ethics, 165*, 45–66.
- Leo, X., & Huh, Y. E. (2020). Who gets the blame for service failures? Attribution of responsibility toward robot versus human service providers and service firms. *Computers in Human Behavior, 113*, 106520.
- Li, J. J., Bonn, M. A., & Ye, B. H. (2019). Hotel employee's artificial intelligence and robotics awareness and its impact on turnover intention: The moderating roles of perceived organizational support and competitive psychological climate. *Tourism Management, 73*, 172–181.
- Li, J., & Huang, J. S. (2020). Dimensions of artificial intelligence anxiety based on the integrated fear acquisition theory. *Technology in Society, 63*, 101410.
- Lin, J., Panganiban, A. R., Matthews, G., Gibbins, K., Ankeney, E., See, C., ... Long, M. (2022). Trust in the danger zone: Individual differences in confidence in robot threat assessments. *Frontiers in Psychology, 13*, 601523.
- Lucas, T., Rudolph, C., Zhdanova, L., Barkho, E., & Weidner, N. (2014). Distributive justice for others, collective angst, and support for exclusion of immigrants. *Political Psychology, 35*(6), 775–793.
- Mann, F. D., Krueger, R. F., & Vohs, K. D. (2020). Personal economic anxiety in response to COVID-19. *Personality and Individual Differences, 167*, 110233.
- Mathias, P. (2013). *The first industrial nation: The economic history of Britain 1700–1914*. London: Routledge.
- McClure, P. K. (2018). “You're fired,” says the robot: The rise of automation in the workplace, technophobes, and fears of unemployment. *Social Science Computer Review, 36*(2), 139–156.
- McFarland, S., Webb, M., & Brown, D. (2012). All humanity is my ingroup: A measure and studies of identification with all humanity. *Journal of Personality and Social Psychology, 103*(5), 830–853.
- Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass, 4*(5), 344–357.
- Miao, X., Liu, L., Dang, J., Wei, C., Huang, L., & Liu, Z. (2023). Unity or estrangement under crises? Perceived resource scarcity moderates the effect of a common threat on intergroup cooperation. *Social Psychological and Personality Science, 15*(5), 659–669.
- Miloyan, B., Bienvenu, O. J., Brilot, B., & Eaton, W. W. (2018). Adverse life events and the onset of anxiety disorders. *Psychiatry Research, 259*, 488–492.
- Neufeind, M., O'Reilly, J., & Ranft, F. (2018). *Work in the digital age: Challenges of the fourth industrial revolution*. London: Rowman and Littlefield.
- Osgood, J. M., & Muraven, M. (2015). Self-control depletion does not diminish attitudes about being prosocial but does diminish prosocial behaviors. *Basic and Applied Social Psychology, 37*(1), 68–80.
- Peng, K., Nisbett, R. E., & Wong, N. Y. C. (1997). Validity problems comparing values across cultures and possible solutions. *Psychological Methods, 2*(4), 329–344.
- Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology, 56*, 365–392.
- Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review, 10*(4), 336–353.
- Roccas, S., & Amit, A. (2011). Group heterogeneity and tolerance: The moderating role of conservation values. *Journal of Experimental Social Psychology, 47*(5), 898–907.
- Römpke, A. K., Fritsche, I., & Reese, G. (2019). Get together, feel together, act together: International personal contact increases identification with humanity and global collective action. *Journal of Theoretical Social Psychology, 3*(1), 35–48.
- Rughiniş, C., Zamfirescu, R., Neagoe, A., & Rughiniş, R.

- (2018, April). Visions of robots, networks and artificial intelligence: Europeans' attitudes towards digitisation and automation in daily life. In *The international scientific conference eLearning and software for education* (Vol. 2, pp. 114–119), Bucharest, Romania.
- Sherif, M. (1966). *In common predicament: Social psychology of intergroup conflict and cooperation*. New York: Houghton Mifflin.
- Shoss, M. K., & Ciarlante, K. (2022). Are robots/AI viewed as more of a workforce threat in unequal societies? Evidence from the eurobarometer survey. *Technology, Mind, and Behavior*, 3(2), 1–13.
- Sidanius, J. (1993). The psychology of group conflict and the dynamics of oppression: A social dominance perspective. In S. Iyengar & W. McGuire (Eds.), *Explorations in political psychology* (pp. 183–219). Durham, NC: Duke University Press.
- Simpson, B., & Willer, R. (2015). Beyond altruism: Sociological foundations of cooperation and prosocial behavior. *Annual Review of Sociology*, 41(1), 43–63.
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, 89(6), 845–851.
- Stephan, W. G., Ybarra, O., & Morrison, K. R. (2009). Intergroup threat theory. In T. D. Nelson (Ed.), *Handbook of prejudice* (pp. 43–59). Mahwah, NJ: Lawrence Erlbaum.
- Sumner, W. G. (1906). *Folkways: A study of mores, manners, customs and morals*. Mineola, NY: Dover Publications.
- Susskind, R. E., & Susskind, D. (2015). *The future of the professions: How technology will transform the work of human experts*. New York: Oxford University Press.
- Tabri, N., Wohl, M. J. A., & Caouette, J. (2018). Will we be harmed, will it be severe, can we protect ourselves? Threat appraisals predict collective angst (and its consequences). *European Journal of Social Psychology*, 48(1), 72–85.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of Intergroup Relations* (pp. 7–24). Chicago, IL: Nelson-Hall.
- Touré-Tillery, M., & Light, A. E. (2018). No self to spare: How the cognitive structure of the self influences moral behavior. *Organizational Behavior and Human Decision Processes*, 147, 48–64.
- Uenal, F., Sidanius, J., Roozenbeek, J., & Linden, S. (2021). Climate change threats increase modern racism as a function of social dominance orientation and ingroup identification. *Journal of Experimental Social Psychology*, 97, 104228.
- Vezzali, L., Andrighetto, L., Drury, J., Di Bernardo, G. A., & Cadamuro, A. (2017). In the aftermath of natural disasters: Fostering helping towards outgroup victims. In H. Zagefka & E. Van Leeuwen (Eds.), *Intergroup helping: The positive side of intergroup behaviour* (pp. 305–330). New York, NY: Springer.
- Vu, H. T., & Lim, J. (2021). Effects of country and individual factors on public acceptance of artificial intelligence and robotics technologies: A multilevel SEM analysis of 28-country survey data. *Behaviour and Information Technology*, 41(7), 1515–1528.
- Wohl, M. J. A., & Branscombe, N. R. (2008). Collective angst: How threats to the future vitality of the ingroup shape intergroup emotion. In H. A. Wayment & J. J. Bauer (Eds.), *Transcending self-interest: psychological explorations of the quiet ego* (pp. 171–181). Washington, DC: American Psychological Association.
- Wohl, M. J. A., Branscombe, N. R., & Reysen, S. (2010). Perceiving your group's future to be in jeopardy: Extinction threat induces collective angst and the desire to strengthen the ingroup. *Personality and Social Psychology Bulletin*, 36(7), 898–910.
- Wohl, M. J. A., Giguère, B., Branscombe, N. R., & McVicar, D. N. (2011). One day we might be no more: Collective angst and protective action from potential distinctiveness loss. *European Journal of Social Psychology*, 41(3), 289–300.
- Wohl, M. J. A., Squires, E. C., & Caouette, J. (2012). We were, we are, will we be? The social psychology of collective angst. *Social and Personality Psychology Compass*, 6(5), 379–391.
- World Bank. (2011). *World Bank national accounts data, and OECD National Accounts data files*. Retrieved August 27, 2024, from <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2011&start=2011>
- World Bank. (2016). *World Bank national accounts data, and OECD National Accounts data files*. Retrieved August 27, 2024, from <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2016&start=2016>
- Wu, N. (2022a). Misattributed blame? Attitudes toward globalization in the age of automation. *Political Science Research and Methods*, 10(3), 470–487.
- Wu, N. (2022b). “Restrict foreigners, not robots”: Partisan responses to automation threat. *Economics & Politics*, 35(2), 1–24.
- Wu, T. J., Liang, Y., & Wang, Y. (2024). The buffering role of workplace mindfulness: How job insecurity of human-artificial intelligence collaboration impacts employees' work-life-related outcomes. *Journal of Business and Psychology*, 39, 1395–1411.
- Xu, L., Wang, X., Yu, F., & Peng, K. (2024). The influence of perceived robot threat on workplace objectification. *Acta Psychologica Sinica*, 56(2), 210–225.
- [许丽颖, 王学辉, 喻丰, 彭凯平. (2024). 感知机器人威胁对职场物化的影响. *心理学报*, 56(2), 210–225.]
- Yang, Y., Sedikides, C., Wang, Y., & Cai, H. (2024). Nature nurtures authenticity: Mechanisms and consequences. *Journal of Personality and Social Psychology*, 126(1), 79–104.
- Yam, K. C., Tang, P. M., Jackson, J. C., Su, R., & Gray, K. (2023). The rise of robots increases job insecurity and maladaptive workplace behaviors: Multimethod evidence. *Journal of Applied Psychology*, 108(5), 850–870.
- Yogeeswaran, K., Zlotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., & Ishiguro, H. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *Journal of Human-Robot Interaction*, 5(2), 29–47.
- Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. In N. Bostrom, & M. M. Čirković (Eds.), *Global catastrophic risks* (Vol. 1, p. 184). New York: Oxford University Press.
- Zlotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, 100, 48–54.

## Perceived robot threats reduce pro-social tendencies

XU Liying, ZHANG Yuyan, YU Feng

*(Department of Psychology, Wuhan University, Wuhan 430072, China)*

### Abstract

The advent of artificial intelligence (AI) presents an inevitable challenge to human society, raising concerns that range from the displacement of human jobs by robots to the increasingly ambiguous boundaries between humans and machines. This phenomenon is conceptualized as the realistic and symbolic threat posed by robots, which engenders collective angst among human beings. This collective angst reflects a concern for the future of the group, manifesting in negative emotions among group members apprehensive about their future. When a group perceives itself to be in crisis, societal cohesion may weaken, leading to increased divisions and aggression, and consequently, a decline in pro-social tendencies. Importantly, this decline is typically observed only toward out-group members, whereas the perceived robotic threat does not diminish pro-social behavior toward in-group members. Additionally, moral comparisons serve as a moderating factor in this dynamic. Downward moral comparisons, which enhance self-improvement and create a moral license that justifies both moral and immoral behaviors (e.g., a reduction in pro-social behaviors), suggest that the effect of robotic threat on pro-social tendencies is more pronounced among individuals who engage in such comparisons.

This research consists of eight studies based on the basic hypothesis that perceived robotic threat reduces people's pro-social tendencies, with collective angst acting as a mediating mechanism, and inside-outside group and moral comparison tendencies moderating its effects. Studies 1a and 1b used pre-existing archival databases to calculate the relationship between perceived robot threat at the individual level and pro-social behavior at the national level. Study 2, on the other hand, used a questionnaire to examine collective angst as a mediating variable between robot threat and reduced pro-social tendencies. Study 3 experimentally manipulated robot threat to similarly validate the mediating role of collective angst. Study 4 excluded the influence of other potential mediators. Studies 5 and 6 also used experimental methods to respectively explore the moderating role of internal and external groups and moral comparison tendencies, and to clarify the boundary conditions between perceived robot threat reducing pro-social tendencies. Study 7 further validated the influence of perceived robot threat on prosocial behavior through behavioral observation.

The main findings of this research are as follows. First, perceived robotic threat indeed reduces people's pro-social tendencies, even at the group level (Studies 1~7); Second, the mechanism is mediated by collective angst, that is, perceived robotic threat increases collective anxiety, which reduces pro-social tendencies in order to alleviate anxiety (Studies 2~4); Third, This effect is moderated by in-group and out-group distinction, whereas perceived robotic threat primarily reduces pro-social tendencies for out-group members (Study 5); Fourth, this effect is moderated by the tendency to make moral comparisons, that is, perceived robot threat mainly reduces pro-social tendencies for downward moral comparators (Study 6).

As findings on the social consequences of robotic threats in prior research have been inconsistent, the results of this research strongly support the conclusion that robotic threats result in negative rather than positive interpersonal outcomes. The reason for this may be the difference between the internal and external groups that people psychologically delineate when they are faced with specific people to engage in interpersonal behavior after experiencing a robotic threat. Our research theoretically explains why robotic threats produce negative interpersonal outcomes.

**Keywords** pro-social tendencies, perceived robot threat, collective angst, out-groups members, moral comparisons