

算法决策对员工公平感的影响研究

景 怡 邱凌云 任 润*

摘 要：随着人工智能技术的快速发展，算法正逐步替代人类管理者，广泛用于组织管理决策。不过，现有文献在算法（相对于人类管理者）决策对员工公平感知的影响上尚未得到一致结论。本文基于刻板印象内容模型，探索了在不同的组织管理场景下，员工对这两种不同决策主体的公平感知差异及其内在机制。情景实验结果显示，在面对不利结果时，人们普遍认为算法比人类管理者更为公平，并且这一现象背后的影响机制受到任务类型的调节：对于低主观性任务，算法被认为更公平的原因是人们认为人类管理者（相比算法）的能力更差；对于高主观性任务，人们则认为人类管理者（相比算法）更为冷漠。本文拓展了算法公平的相关文献，为探究算法决策对组织管理的影响提供了新的视角。

关键词：算法决策；公平感；刻板印象内容模型；任务类型

中图分类号：F270

JEL 分类号：M10

一、引 言

伴随着新一代人工智能（Artificial Intelligence，简称 AI）技术的发展和普及，算法决策（Algorithmic Decision-Making）被企业广泛应用于组织管理中（Lee, 2018; Daugherty et al., 2019; Araujo et al., 2020; Newman et al., 2020）。例如，一项近期的调查显示，有超过三成的受访企业在人力资源管理中已经采用了基于人工智能的算法决策，应用场景涉及考勤、员工招募、薪酬与福利、行政事务等（李超平和徐世勇, 2019）。

相较于人类管理者，算法在工作效率上具有明显优势（Lee, 2018; Leicht-Deobald et al., 2019; Lewis and Marc, 2019）。例如，在外卖行业中应用算法分派订单，可以显著提升配送人员的服务效率（Kim et al., 2021; Wu et al., 2023）；在销售电话外呼场景中，基于人工智能算法的聊天机器人比经验不足的员工的效率高出 4

* 景怡，中国工商银行博士后科研工作站，E-mail: yi.jing@icbc.com.cn；邱凌云（通信作者），北京大学光华管理学院，E-mail: qiu@gsm.pku.edu.cn；任润，北京大学光华管理学院，E-mail: renr@gsm.pku.edu.cn。作者感谢国家自然科学基金面上项目（72272004）对本文研究的资助。作者感谢匿名审稿人和编辑部的宝贵意见，当然文责自负。

倍,达到熟练员工的水平(Luo et al., 2019)。

尽管如此,员工如何感知算法决策的公平性依然是组织亟须回答的问题。作为组织成员对自身工作环境的重要心理感知之一(Greenberg, 1990),公平感会显著影响员工的工作满意度、组织承诺、离职意愿、任务表现和组织公民行为(Ambrose and Schminke, 2009; Choi, 2011)。例如,当员工认为自己(或他人)受到不公平对待时,不仅会感到压力增加(Elovainio et al., 2001),还会因减少工作投入而导致绩效降低(Lau and Oger, 2012),甚至引发偏差(Colquitt et al., 2013)和破坏性行为(Skarlicki and Folger, 1997)。

在决策主体由人类转变为算法的过程中,员工如何感知和评价算法决策的公平性会显著影响组织数字化转型的效果(裴嘉良等, 2021)。如果员工认为算法决策不如人类决策公平,他们可能会因此对算法管理加以抵制,甚至做出不利于企业的行为(Wu et al., 2023)。然而,现有研究在算法决策对公平感的影响上尚未取得一致结论。有研究发现,由于算法决策大多基于数据和数学模型,人们相信算法可以有效减少由人类决策者主观偏差而引发的不公平感(Miller and Keiser, 2021)。例如,一项有关视频面试的研究发现,申请人认为AI面试在程序公平性、客观性和一致性等方面优于由人类面试官根据直觉和经验做决定的传统评估方式(Kim and Heo, 2022)。但也有研究发现,由于算法无法考虑某些非量化的信息、缺乏对个体差异的关注或缺乏足够的透明度(Lee et al., 2015),也有人认为算法决策不如人类决策公平(Newman et al., 2020)。

不过,这些研究中的大多数都是在不考虑决策结果的前提下探讨人们对算法决策的一般性态度。但近期的研究发现,不同的决策结果会显著调节人们对算法的公平性感知。例如,魏昕等(2021)发现,当决策结果对员工有利时,决策者类型(人工 vs. 算法)不会对员工的公平感产生显著影响;但当决策结果对员工不利时,他们会觉得人工决策更不公平。因此,有必要根据不同的决策结果深入探索人们对算法决策和人类决策的相对偏好。

为深入探索员工对算法决策的公平感知,本文聚焦于员工在面对不利结果时对不同决策者(AI vs. 管理者)的公平感差异及其形成机制。在现实情境中,人们往往在遇到不利于自身的决策结果时更容易诱发不公平感。这是因为公平感在很大程度上受到员工对决策结果归因的影响(Ployhart and Ryan, 1997)。自我服务偏差理论(Self-Serving Bias Theory)指出,在面对有利结果时,人们更倾向于做内部归因,即将结果归结为自身原因;而在面对不利结果时,人们更倾向于做外部归因,即将结果归因到包括决策者在内的各种外部因素(Heider, 1958; Campbell and Sedikides, 1999)。

为解释不同决策主体对员工公平感的影响机制,本文借助Fiske et al. (2002)提出的刻板印象内容模型(Stereotype Content Model, 简称SCM),提出将员工对决策

者的感知能力和感知热情作为中介机制。刻板印象内容模型指出,人们会从能力和热情两个维度认知和评价他人或群体(Judd et al., 2005)。这两种维度的划分源于人类在进化中演化出的一种策略,即感知者想要知道他人积极或消极的意图,以及他们有效实现这一意图的能力。具体而言,能力维度包含与感知能力有关的特质,包括人的智力、技能、创造力和效能等;热情维度包含与感知意图相关的特质,包括人的友好、利他、真诚、可信和道德等(Fiske et al., 2007)。尽管该模型传统上被用于对人的评价,但近年来的研究也发现,人们在与技术产品互动时,也会基于感知能力和感知热情形成对产品的印象(McKee et al., 2023)。例如,感知能力能够反映人们对机器人是否可以准确可靠地执行任务的判断,感知热情能够反映人们对机器人是否抱有善意的判断(Belanche et al., 2021),感知能力和感知热情都会影响人们对机器人的信任、使用意愿和满意度(van Doorn et al., 2017; Gelbrich et al., 2021)。因此,本文提出,人们也会从能力和热情两个维度对算法决策者进行评价,进而影响其公平感的形成。

此外,本研究还探讨了任务类型对上述中介机制的调节作用。已有研究表明,任务类型会显著调节人们对算法或人类决策者的信任、使用意愿或偏好。因此,现有文献中不一致的研究发现也可能与决策任务的类型和场景有关(蒋路远等, 2022)。例如,在面对约会对象的推荐广告时,消费者点击算法建议的次数显著少于人类建议;而当面对财务建议推荐时,两者在点击次数上则没有差异(Castelo et al., 2019)。在相关文献中,研究者从不同维度对任务类型进行划分,例如主客观性(Castelo et al., 2019; Srinivasan and Sarial-Abi, 2021)、人文或“机械”属性(Lee, 2018)、复杂性和困难性(Xu et al., 2020; Bogert et al., 2021)和交互性(Srinivasan and Sarial-Abi, 2021)等。本文选择了决策任务的主观性——即完成该任务在多大程度上依赖人类的直觉判断和情感能力——作为调节变量。

综上所述,本研究试图在组织管理情境中回答以下三个研究问题:①当面对不利于自己的决策结果时,算法和人类决策者是否会影响员工对组织的公平感知?②员工对算法和人类决策者的感知能力和感知热情是否存在差异?这种差异是否可以解释他们在公平感上的不同?③任务的主客观性是否会对上述的中介机制造成影响?

二、文献综述

(一) 算法管理及其影响

随着人工智能的发展,算法不仅用于提供决策辅助,甚至能够扮演管理者的部分角色。例如,算法决策已经被用于搜索、过滤、选择和推荐求职者(Daugherty et al., 2019)、为员工分配任务(Hoshino et al., 2018)以及对员工进行评估(Dhir and

Chhabra, 2019)等。这种将管理职能委托给算法的做法被称为“算法管理”(Algorithmic Management)(Lee et al., 2015; Lee, 2018)。

算法管理的早期研究主要集中于零工经济情境下对非标准合同工人的管理,关注算法如何对传统的人工监督管理进行替代和补充(Tambe et al., 2019; Newlands, 2021)。近年来,算法管理在其他工作环境中也得到广泛应用,例如算法可以帮助人力资源部门过滤求职者(Leicht-Deobald et al., 2019)、解雇绩效欠佳的仓库工人、提高工作士气等(Gal et al., 2020)。算法管理的相关研究也相应转向工作环境中工人、管理人员和算法之间的互动。其中,不少研究关注了员工对算法管理的感知与行为,探讨了算法管理对员工信任、人际关系、心理契约、自我认知、公平感知等的影响。例如,研究发现员工认为算法比人类领导者在诚信和透明度上的得分更高,而人类领导者则被认为更具适应性和更有善意(Höddinghaus et al., 2021)。还有研究发现,在管理任务中部署算法会导致员工对其他员工的物化,从而降低亲社会动机,减少帮助他人的愿望(Granulo et al., 2024)。算法管理的规范性和一致性可以激励员工维持与公司和雇主的心理契约、增加互惠,但员工也可能由于算法管理减少与其他人的沟通和互动,更容易产生被侵犯的感觉,从而降低与组织的心理契约(Tomprou and Lee, 2019)。算法管理还被发现会导致员工对自我地位的评价降低。如Jago and Carroll(2024)发现,与人类管理相比,人们认为受到算法管理的工作更加简单,这种信号导致人们降低对自我地位的感知并带来显著的负面情绪。在面对算法管理时,员工可能还经历角色冲突和角色模糊(Tarafdar et al., 2023)。

(二)对算法管理的公平感知

算法管理在组织中的广泛应用引发了众多研究问题。其中,员工的公平感知是一个研究者和管理者较为关注的关键问题。公平感是导致员工、客户和个人消费者对算法形成厌恶感的重要原因之一(Haak, 2017)。随着算法管理被逐渐应用在较为敏感和高风险的决策场景,不仅管理者愈加关注算法决策的公平性(De Cremer, 2020),受到决策影响的员工也开始关注算法决策是否会以关心、尊重和同情的态度对待他们,是否能够充分把握人类行为的复杂性和细微差别,以及能否给出公平公正的决策结果(Binns et al., 2018; Kellogg et al., 2020; McGuire and De Cremer, 2023)。

在此背景下,算法公平的相关研究可分为两个研究主题:一部分研究从算法与人类决策者的二元对比出发,探讨个体在面对算法或人类决策主体时在公平感知上的差异;另一部分研究则从算法设计的角度出发,探讨算法作为决策者时,能够影响个体公平感知的相关因素及其作用机制。本文重点关注第一个研究主题的相关文献。

一方面,部分研究发现,在组织中应用算法决策会带来相较于人类决策者更高的公平感知。这是因为,人们通常认为算法决策更加准确一致,不存在人类的主观

偏见和故意。人们倾向于认为,算法在设计上主要基于客观历史数据和事实,遵循固定模型和规则(Lindebaum and Ashraf, 2021),因此能够确保其决策过程对所有接受者都是一致的(Howard et al., 2020),不会受到组织政治等因素的影响(Park et al., 2021)。例如,一项针对外卖配送员的研究表明,当面对不利于自己的工作排班时,员工认为人类领导做出的决策相较于算法做出的决策更不公平(魏昕等, 2021)。在工作分配场景中,采用算法分配会提高员工的公平感知和生产力(Bai et al., 2022)。在招聘场景中,参与者认为高度自动化的面试更加一致(Langer et al., 2019);与人工评估相比,算法不仅更客观,也更不容易受到操纵(Hilliard et al., 2022)。

另一方面,也有部分研究发现算法决策不如人类决策公平(Langer et al., 2019; Zhang and Yench, 2022)。对算法决策的不公平感知的驱动因素主要包括决策过程的不透明(Lee et al., 2015)、较少的社会互动(Langer et al., 2020)、较少的表现机会(Kaibel et al., 2019)、较低的自尊(Zhang and Amos, 2023)等。例如,在人力资源管理中使用算法决策时,员工会认为自己被简化为一个数字,仅代表达到目的的手段(Newman et al., 2020),认为自己没有受到组织的尊重和重视(Zhang and Amos, 2023),从而降低了其公平感知。在招聘场景中,求职者认为算法面试减少了个人的表现机会和复议机会,缺乏双向沟通(Noble et al., 2021)。也有少数研究发现,员工在面对人类管理者和算法管理者时,在绩效、内在动机、公平感知和承诺等方面没有表现出显著差异(Dong et al., 2024)。此外,还有研究从某一视角出发,尝试解释人们对算法决策和人类决策在公平感知上的差异。例如, Li and Bitterly (2024) 基于情感即社会信息(Emotions as Social Information, 简称EASI)模型,提出由于算法无法体验情感、无法同情他人,因此员工难以感知算法的善意,进而使他们降低了对算法管理的信任。

综上所述,现有对于人类决策者与算法在公平感知上的二元对比的研究尚未取得一致结论。这些不一致可能源自不同研究在算法的设计特征、决策结果和应用场景等各个方面都存在较大差异,而个体对于算法的公平感判断很大程度上取决于所处的具体决策情景(蒋路远等, 2022)。例如,大部分已有研究并未给定算法决策或人类决策的具体结果。因此,有必要在更为具体的决策结果和任务类型条件下对算法决策如何影响人们的公平感知进行进一步探究。

三、研究假说

本研究提出,当面临对自己不利的决策时,员工的公平感会在很大程度上受其对决策者行为归因的影响。作为一种对“人们如何得出因果推论,他们做出什么样的推论,以及这些推论的后果是什么”的理论解释(Folkes, 1988),归因理论指出,

人们为了维持或提高自我价值,通常会将有利结果归因到与自身相关的因素(内部归因),而将不利结果归因为他人或环境等外部因素(外部归因)(Wolosin et al., 1973; Kelley and Michela, 1980; Belanche et al., 2020; Lei and Rau, 2021)。对公正世界的信念(Lerner, 1965)和防御性归因(Shaver, 1970)的相关研究也都发现,当人们感知受到伤害时,他们有强烈的动机将伤害归咎于他人。

当员工对不利于自己的决策结果进行外部归因时,他们会因此形成对于决策者的主观判断。本文借助刻板印象内容模型来描绘这种判断。刻板印象内容模型定义了社会知觉和社会判断的两个基本维度:能力和热情(Fiske et al., 2002)。该理论指出,无论对象是人、群体还是文化(Judd et al., 2005),人们都会从能力和热情这两个维度形成判断。能力维度包括与感知能力相关的特质,包括智力、技能、创造力和效能,而热情维度则包括与感知意图相关的特质,包括友好、乐于助人、真诚、可信和道德等(Fiske et al., 2007)。受到不利决策结果影响的员工也会从决策者的能力和热情出发,对决策者形成认知,进而影响其对公平的判断。

在组织管理中,员工对决策者能力和热情的感知会影响员工的组织认同、工作效率、服从和公平感知等。一方面,员工认为决策者的能力决定了其能否根据准确的信息做出遵循公平原则和一致的决定(Cojuharencu et al., 2017)。当下属对主管的知识、技能和处理任务相关问题的能力有信心时,不仅有助于提高工作效率(Kanfer and Ackerman, 1989),还可以增加其对上级工作指示的服从(Yang et al., 2009)。Mossholder et al. (1998)还发现,人们倾向于相信能力强的决策者更能够权衡不同的观点和利益,做出更深入的分析,从而做出更准确和公正的决定。另一方面,决策者的热情能够增加员工对其的信任和公平感知。热情维度包含的一些属性(如友好、乐于助人、真诚、可信和道德等)与公平相关。Hollensbe et al. (2008)发现,主管的友好和支持正向影响新人对主管的公平看法。同时,对热情的判断也被用来衡量他人的意图,热情的人被认为更有可能追求利他的结果(Fiske et al., 2007),一个热情的决策者可能具有公平动机(Cropanzano, 2012)。此外,热情的领导者还会提升下属的工作动力以及对与领导关系的满意程度(Tjosvold, 1984)。热情的领导者往往具备较强的同理心,能够通过富含情绪信息的人际互动(Daus and Ashkanasy, 2005),营造良好人际氛围,减少员工的不公平感。

相较于对人类决策者的判断,员工对算法的能力和热情的感知有所不同。研究发现,人们对他人能力的判断与对象的分析能力高度相关,如解决数学问题、处理新信息和逻辑推理等(Furnham, 2001)。随着大数据和人工智能技术的快速进步,算法在许多认知任务上的表现已经超过人类。此外,算法在信息处理能力和效率上也显著高于人类,这也增加了人们对算法能力的判断。例如,Fumagalli et al. (2022)发现,在招聘场景中人们通常认为算法能够更快速地做出决定;Choung et al. (2023)发现,在申请工作时,人们认为算法决策更有能力、更值得信赖、更有

用。Tong et al. (2021) 的大规模实地试验表明,在为员工生成绩效反馈时,人工智能系统能够显著提高反馈的准确性和信息分析的一致性,提高了绩效反馈的质量。Kellogg et al. (2020) 也发现,算法可以提供比人和其他传统方法更全面的反馈。算法在数据分析方面的优势在日常工作和生活中被广泛熟知,人们正在逐渐接受一种新的刻板印象,即算法能够更准确高效地完成以计算为基础的认知任务,能处理更为复杂的问题。因此,人们更有可能在面对不利于自己的决策结果时,认为人类决策者比算法更有可能犯错。

与此同时,算法也被认为没有情感和缺乏主观意图,不像人类那样具有歧视性(Fumagalli et al., 2022)。与人类相比,算法被认为具有较少的能动性和情感能力(Gray et al., 2007; Gray and Wegner, 2012)。基于算法的各种系统或代理本质上是由人类制造的服务于人类的非人类机器(Russell and Norvig, 2003),尽管算法代理的行为依然会服务于其设计者设定的目标,但人们通常认为算法不存在个人动机。

在面对不利于自己的决策结果时,缺少意图意味着算法相比人类具有更少的主观恶意。例如,Srinivasan and Sarial-Abi (2021) 发现,算法相较于人类的能动性更低,这会降低算法犯错后的责任感知。宋晓兵和何夏楠(2020)指出,面对不公平的价格决策,消费者更可能认为销售人员蓄意抬高价格,而认为AI系统的决定并非故意。Garvey et al. (2023) 发现,当产品或服务不如预期时,消费者不会因此认为AI代理自私或有恶意;当产品或服务高于预期时,消费者也不会因此认为AI代理心怀善意。也就是说,面对不利结果,人们可能认为算法相比人类更不可能存在自私或其他恶意,这有助于减少人们对决策者的感知冷漠。

综上所述,本文提出假说1:

假说1 当面对不利于自己的决策结果时,员工认为算法决策比人类决策更公平。

已有研究表明,人们对算法或人类决策者的信任、使用意愿或偏好会受到决策任务的主观性的调节(Castelo et al., 2019; Srinivasan and Sarial-Abi, 2021)。主观性是指决策任务受个体直接影响的程度,主观性高的任务受个人心理或个体差异(如感知、情感和经验)的影响较大(Choon and Embi, 2012)。Castelo et al. (2019) 发现,在完成客观任务(如财务建议)时,人们对人类和算法建议者在信任程度没有显著差异,但在完成主观任务(如推荐约会对象)时,人们会更信任人类推荐。Lee (2018) 在组织场景中的研究也发现,一般而言,员工认为算法在完成需要机械技能的客观任务(如工作排班和时间分配)时与人类管理者同样公平,但人类管理者在完成需要人类技能的主观任务(如雇佣决策和绩效评估决策)时更加公平。

不过,现有关于任务主观性对算法态度影响的研究多是在决策结果不确定的前提下探讨人们对算法或人类决策的事前感知或总体态度。当某种决策结果被明确时,任务主观性是否仍会调节人们对算法和人类决策者的感知尚未得到证实。尤其是考虑到公平感的形成过程受情境因素的影响很大(Buckley, 2012),因此,有必要在具

体的决策结果下探究任务主观性的调节作用。

本研究提出,任务主观性会影响人们对决策者能力和热情的权重判断。一方面,对于低主观性任务,人们认为决策者在完成任务时更多地依赖客观事实和标准,较少依靠自身的主观判断和意图,结果主要由决策者能否准确有效地按照标准和规则、基于可量化和可验证的信息进行决策来决定,即决策者的能力因素。Sun et al. (2023) 的研究表明,在依靠虚拟助手执行诸如搜索任务在内的功能性任务时,用户主要依据能力(而非热情)来对虚拟助手进行评价。另一方面,领导力相关文献指出,对领导者的能力和热情的感知取决于员工期望和现实的匹配(Lambert et al., 2012)。因此,在低主观性任务中,由于决策结果多由领导者难以控制和影响的客观因素决定,员工并不期望与管理者建立情感联系会有助于获得对自己有利的结果。此时管理者的情感能力(即热情)的重要性相对较低。Li and Bitterly (2024) 的研究也发现,在对同理心需求较低的场景中,员工对 AI 管理者和人类管理者在感知善意上不存在显著差异。相比之下,人们普遍存在算法(相较于人类)在完成客观任务时的能力更强这一外行信念(von Walter et al., 2022)。因此,在低主观性任务中,人们会更容易将人类决策者造成的不利结果归咎为决策者的能力不足,进而产生较强的不公平感。此时,感知无能的中介作用会显著高于感知冷漠的中介作用。

在面对高主观性任务时,人们则更倾向于将决策结果归因为决策者的主观意图而非能力(Brockner et al., 2007)。因此,不利的决策结果往往被员工视为决策者故意释放的信号(Voußem et al., 2016)。由于人们普遍存在算法不具有主观恶意这一外行信念(Srinivasan and Sarial-Abi, 2021; Garvey et al., 2023),因此,在高主观性任务中,人们会更容易将不利结果归咎为人类决策者的主观恶意(即冷漠),进而产生较强的不公平感。此时,感知冷漠的中介作用会显著高于感知无能的中介作用。

综上所述,本文提出假说2和假说3:

假说2 在低主观性任务中,当面对不利的决策结果时,决策主体对员工公平感的影响主要被感知决策主体的能力(而非热情)中介。

假说3 在高主观性任务中,当面对不利的决策结果时,决策主体对员工公平感的影响主要被感知决策主体的热情(而非能力)中介。

四、研究方法

主实验采取2(决策主体:人 vs. 算法)×2(任务主观性:低 vs. 高)的组间实验设计。

(一) 实验设计

为操控决策任务的主观性,本研究选取了工作排班和绩效评估分别作为低主观性任务和高主观性任务。这不仅因为这些任务在组织管理中较为常见,还因为算法已被广泛应用于完成这些任务 (Lee, 2018)。现有研究发现,工作排班主要依靠人的计算能力,需要处理定量数据,管理者在决策时相对客观 (Lee, 2018),而绩效评估则涉及对他人的判断 (Milkovich and Wigdora, 1991),更容易受到管理者个人经验和主观偏好的影响 (Zhang and Amos, 2023)。为确保操控的有效性,我们设计了一个前测实验来验证人们对这两种任务在感知决策主观性上的差异,并且排除任务复杂性的替代解释 (Bogert et al., 2021)。

我们通过 Credamo 平台招募了 35 名被试。前测采取组内设计,每位被试需要分别阅读对工作排班和绩效评估的任务说明,两个任务呈现的次序随机。针对工作排班任务,被试被要求想象自己是一家连锁快餐店的店长,需要根据快餐店的客流量为管理的 10 名店员安排一周的工作排班,在保证所有店员的工作时长基本一致的同时,需要把他们公平地分配到工作相对轻松和繁忙的时段。在绩效评估场景中,被试被要求想象自己是一家电商平台的电话客服中心主管,需要对管理的 10 名客服人员每周的工作表现进行考核。主管需要随机抽取每位客服人员的部分电话录音进行检查,并根据客服人员接电话时的语气、语调、态度积极程度对其工作表现给出评价 (任务描述详见附录表 A1)。

在阅读完任务说明之后,被试回答对每个任务的主观性感知和人口统计学信息。对任务主观性的测量为单一题项 (7 点量表),被试回答“您觉得完成这个任务更多依靠可量化的客观信息还是个人的主观判断或直觉?”,其中 1 = 完全依靠客观信息,7 = 完全依靠主观判断。对任务复杂性的测量为单一题项 (7 点量表),被试回答“您觉得完成这个任务有多复杂?”,其中 1 = 一点也不复杂,7 = 非常复杂。前测结果显示,绩效评估任务比工作排班任务的感知主观性更高 ($M_{\text{评估}} = 4.91$, $M_{\text{排班}} = 3.54$, $F(1, 68) = 15.181$, $p < 0.001$),但两者在复杂性上没有显著差异 ($M_{\text{评估}} = 4.89$, $M_{\text{排班}} = 4.94$, $F(1, 68) = 0.026$, $p = 0.872$)。

(二) 测量

本研究采用的量表均来自成熟量表。实验采用了反向量表,即分别测量被试对不公平、决策者无能和决策者冷漠的感知。这主要基于以下考虑:①本研究主要关注员工的不公平感,使用反向量表与实验目的相一致;②使用反向量表能够更有效地避免默许偏差 (Acquiescent Response Bias),即被试更倾向于对量表中的题项做出积极评价 (Jackson and Messick, 1957)。对感知不公平的测量改编自 Ambrose and Schminke (2009),包括三个题项:“总体而言,您觉得这家公司对待您有多不公

平?”“您觉得自己在多大程度上受到这家公司的不公平对待?”,以及一项反向测量“您认为这家公司在多大程度上做到了公平?”对感知无能的测量改编自 Cuddy et al. (2007),包括三个题项:“总体而言,您觉得【您的经理/算法】在多大程度上符合以下描述:无能/不胜任/愚蠢?”对感知冷漠的测量改编自 Kervyn et al. (2015),包括三个题项:“总体而言,您觉得【您的经理/算法】在多大程度上符合以下描述:不友好/恶意/不可信?”所有题项均采用7点量表。

(三) 实验样本和流程

我们通过 Credamo 平台招募被试,共有 286 人参与实验。在去掉 27 个未通过操控检验或注意力检验的样本后,共得到 259 个有效样本,其中女性占比 69.5%,样本平均年龄 26.64 岁。

被试被随机分配至 4 个实验组中的一组,然后被要求阅读相应实验组的场景说明。在工作排班场景中,被试被要求想象自己在一家快餐店打工。快餐店在不同时段的小时工资相同,但高峰时段的劳动强度要显著高于非高峰时段。被试得知自己每周被安排在高峰时段的次数要明显多于其他同事。作为对决策主体的操控,人类决策组的被试被告知这家店由店长分配每周的排班,算法决策组的被试被告知这家店由一套基于人工智能算法的自动化排班软件自动分配每周的排班。在绩效评估场景中,被试被要求想象自己在一家电商平台公司的客服中心工作,主要负责接听顾客咨询和售后电话。被试得知上个月,尽管自己和大部分同事在各方面的工作表现都比较一致,但得到的绩效评估结果却低于大部分同事。作为对决策主体的操作,人类决策组的被试被告知评估由客服中心的经理完成,算法决策组的被试被告知评估由 AI 算法系统完成(对实验任务的描述详见附录表 A2)。

在阅读完场景说明后,被试首先报告他们对组织的不公平感知、对决策主体无能和冷漠的感知,之后完成操控检验和人口统计学的相关题项。

五、结果分析

(一) 信效度检验

为检验量表的信度和效度,我们计算了每个构念的 Cronbach's Alpha、组合信度(Composite Reliability,简称 CR)、平均方差提取(Average Variance Extracted,简称 AVE)和每个测量项的标准载荷。如表 1 所示,所有构念的 Cronbach's Alpha 和组合信度(CR)均大于 0.8,表明量表具有较好的构念信度;所有构念的平均方差提取(AVE)均大于 0.7,所有测量项的标准载荷均大于 0.8,表明量表具有较好的构念效度。

表 1 构念信度和效度检验

构念	题项	载荷	Cronbach's Alpha	CR	AVE	1	2	3
感知不公平	OJ1	0.802	0.919	0.916	0.786	0.886 [^]		
	OJ2	0.941						
	OJ3	0.910						
感知无能	COM1	0.861	0.887	0.885	0.720	0.554 *	0.633 *	
	COM2	0.879						
	COM3	0.803						
感知冷漠	WAR1	0.929	0.905	0.923	0.799	0.474 *	0.894 [^]	0.849 [^]
	WAR2	0.869						
	WAR3	0.883						

注：* 代表相关系数；[^] 代表对应构念的 AVE 平方根；N = 259。

（二）描述性统计

表 2 展示了核心构念在工作排班和绩效评估场景中的描述性统计结果。

表 2 核心构念的描述性统计结果

任务场景	决策主体	样本量	组织不公平		无能		冷漠	
			均值	标准差	均值	标准差	均值	标准差
工作排班	人类	68	5.85	0.769	5.06	0.948	5.56	1.043
	算法	68	5.41	1.172	4.52	1.561	5.17	1.492
绩效评估	人类	57	5.42	0.943	4.59	1.409	5.24	1.192
	算法	66	4.99	1.190	4.65	1.507	4.64	1.519

注：本表展示了主要变量的汇总统计数据。

（三）方差分析

我们首先对假说 1 加以检验。结果显示，被试总体上认为人类决策比算法决策更不公平（ $M_{\text{人类}} = 5.65$ ， $M_{\text{算法}} = 5.20$ ， $F(2, 255) = 11.138$ ， $p = 0.001$ ）。如图 1 所示，在工作排班场景下，人类管理者造成的不公平感要显著高于算法（ $M_{\text{人类}} = 5.85$ ， $M_{\text{算法}} = 5.41$ ， $F(1, 128) = 7.25$ ， $p = 0.007$ ）。类似的结果也发生在绩效评估场景下（ $M_{\text{人类}} = 5.42$ ， $M_{\text{算法}} = 4.99$ ， $F(1, 121) = 4.605$ ， $p = 0.034$ ）。假说 1 在两种决策任务中均得到支持。

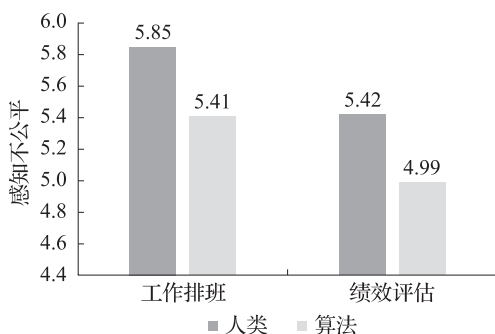


图1 决策者类型对感知不公平的影响

注: $N = 259$ 。

随后,我们分别以感知无能和感知冷漠为因变量进一步进行单因素方差分析。针对感知无能的分析结果见图2。在工作排班场景中,人类决策者的感知无能要显著高于算法 ($M_{\text{人类}} = 5.06$, $M_{\text{算法}} = 4.52$, $F(1, 128) = 4.953$, $p = 0.028$);而在绩效评估场景中,两者之间的差异不显著 ($M_{\text{人类}} = 4.59$, $M_{\text{算法}} = 4.65$, $F(1, 121) = 0.045$, $p = 0.833$)。针对感知冷漠的分析结果见图3。在工作排班场景中,人类决策者与算法的差异不显著 ($M_{\text{人类}} = 5.56$, $M_{\text{算法}} = 5.17$, $F(1, 128) = 1.948$, $p = 0.165$);而在绩效评估场景中,人类决策者的感知冷漠要显著高于算法 ($M_{\text{人类}} = 5.24$, $M_{\text{算法}} = 4.64$, $F(1, 121) = 5.771$, $p = 0.018$)。

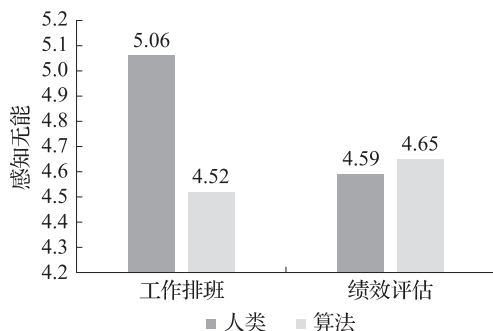


图2 决策者类型对感知无能的影响

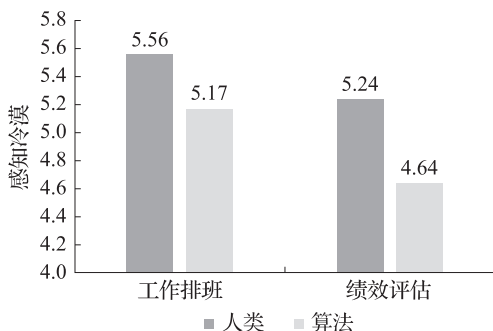
注: $N = 259$ 。

图3 决策者类型对感知冷漠的影响

注: $N = 259$ 。

以上结果表明,在低主观性任务中,决策主体在能力上的感知差异显著高于在热情上的感知差异,换言之,员工此时更倾向于从能力维度比较人类决策和算法决策;而在高主观性的任务中,决策主体在热情上的感知差异显著高于在能力上的感知差异,换言之,员工此时更倾向于从热情维度比较人类决策和算法决策。

(四) 中介检验

我们使用 Hayes (2013) 的 Bootstrapping 方法分别在两种任务场景中检验了感知

无能 and 感知冷漠的中介作用。以决策主体为自变量, 以感知不公平为因变量, 以感知无能和感知冷漠为中介变量, 选择中介模型 Model 4 对假说 2 和假说 3 进行检验。

针对工作排班场景的中介效应检验结果见图 4。当面对不利的排班结果时, 感知无能中介作用显著 (间接效应_{感知无能} = -0.135 8, 95% Boot CI = [-0.283 4, -0.021 7]), 而感知冷漠的中介作用不显著 (间接效应_{感知冷漠} = -0.074 3, 95% Boot CI = [-0.216 7, 0.009 7])。对于低主观性低任务, 算法主要通过降低员工对决策者的感知无能减少对组织的不公平感知。因此, 假说 2 得到支持。

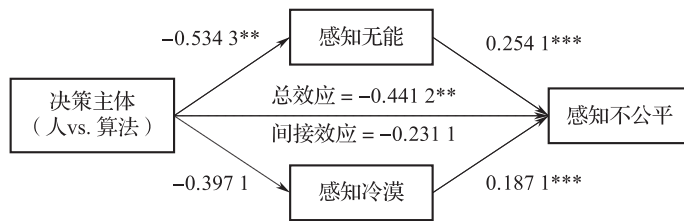


图 4 工作排班场景中中介分析

注: $N=136$; 决策主体: 0 = 人类, 1 = 算法; ** 表示 $p < 0.01$, *** 表示 $p < 0.001$ 。

针对绩效评估场景的中介效应检验结果见图 5。当面对不利的评估结果时, 感知无能中介作用不显著 (间接效应_{感知无能} = 0.015 1, 95% Boot CI = [-0.134 3, 0.166 9]), 而感知冷漠的中介作用显著 (间接效应_{感知冷漠} = -0.174 8, 95% Boot CI = [-0.378 6, -0.024 6])。对于高主观性任务, 算法主要通过降低员工对决策者的感知冷漠减少对组织的不公平感知。因此, 假说 3 得到支持。

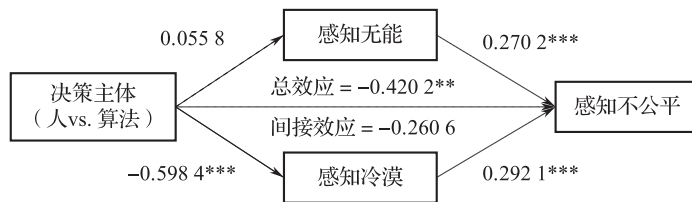


图 5 绩效评估场景中中介分析

注: $N=123$; 决策主体: 0 = 人类, 1 = 算法; ** 表示 $p < 0.01$, *** 表示 $p < 0.001$ 。

六、结论与展望

本文探讨了在组织管理场景中, 尤其是当员工面对不利于自身的决策结果时, 算法 (相较于人类) 决策对员工公平感的影响。研究发现, 当面对不利结果时, 员工总体上觉得采用算法决策的组织比依靠人类决策的组织更为公平, 且这一结论不受任务类型的影响。不过, 算法决策对公平感的影响机制却受到任务类型的调节: 在低主观性任务中, 算法决策通过减少对决策者感知能力的下降增加员工的公平感;

而在高主观性任务中,算法决策则是通过减少对决策者感知热情的下降增加员工的公平感。

(一) 理论贡献

本研究拓展了算法管理和算法公平领域的相关文献,其理论贡献主要体现在三个方面:

首先,关于算法公平的现有文献多数关注人们对算法决策的一般性态度,较少根据不同的决策结果深入分析人们对算法决策的态度(魏昕等,2021)。这可能是现有文献得出不一致的研究结果的原因之一。例如,当决策结果不明时,员工可能会从自身对人类和算法决策者相应属性的预期出发,形成对算法公平性的判断。再如,算法既可能由于基于计算和缺乏同理心等原因而被认为比人类管理者更不公平,也可能由于准确无偏而被认为比人类管理者更公平。本研究弥补了这一不足,在特定的决策结果下探讨员工对于算法决策的态度及其影响机制。

其次,本研究将感知能力和感知热情作为算法决策对公平感影响的作用机制。基于刻板印象内容模型,本研究提出员工在将不利的决策结果归因到决策者后,会从决策者的能力和热情两个维度对决策者进行评价,进而影响对组织公平的感知。这种判断过程不仅适用于人类对象,也适用于基于算法的软件系统。具体而言,由于算法能够收集海量信息并通过精确计算得出结论,人们倾向于认为算法在决策的客观性和准确性上具有优势,比人类决策者的犯错概率更低;同时,由于算法不具备个人意图,人们也倾向于认为算法不会像人类决策者那样怀有主观恶意。对能力和热情的双重判断,最终促使员工认为算法决策更为公平。与采用单一解释机制的文献(Li and Bitterly, 2024)相比,本研究提出的“能力-热情”双路径机制不仅在理论上更为丰富,还能够适应更广泛的研究场景。

最后,本研究拓展了对算法公平的边界条件的探索,验证了任务类型的调节作用。有研究发现,决策任务的类型能够显著调节人们对算法的态度(蒋路远等,2022),本研究则进一步拓展了这一结论。我们提出,任务类型的调节作用不一定体现在人们对算法态度的主效应上,还有可能体现在对公平感的影响机制上。

(二) 实践启示

Sanders and Wood (2020)指出,组织对人工智能无法采用“即插即用”模式,需要充分考虑其在人性和道德层面的影响。随着算法决策在更为广泛的领域和场景下得到部署和使用,组织必须充分考虑算法决策对员工的公平感和道德感的影响。在组织中,员工通常会面临一些对自身不利的决策结果。如果员工将这些结果归因为组织的不公平对待,无疑会在很大程度上伤害员工的士气和对组织的评价。为避免这一结果,组织有必要充分评估和减少员工在面对不利结果时的不公

平感。本研究结果表明,当组织需要做一些员工难以接受的决策时,例如无法在排班时做到绝对公平,或必须对部分员工给出低于同事的绩效评级时,管理者可以考虑使用算法进行决策,以减少员工的不公平感,从而减少员工的消极态度或报复性行为。

与此同时,企业还可以根据决策任务的类型对算法决策进行调整,从而进一步放大算法决策在组织公平方面的优势。具体而言,对于主观性较高的任务,如绩效评估、招聘等,管理者可以在采用算法决策时向员工明确传达算法在设计中没有偏见、不存在负面情绪等特点,例如以更为透明的方式解释算法的设计原则和运作方式,进而放大人们对算法的热情感知并以此增加公平感;针对主观性较低的任务,如生产计划排班等,管理者应当向员工强调算法在准确性和一致性上的优势,例如提供对算法准确性的说明和示例、描述算法使用的大量数据和丰富的信息来源,进而放大人们对算法的能力感知并以此增加公平感。

(三) 局限与展望

本研究还存在一些不足之处,有待后续研究加以弥补和扩展。

第一,本研究中采用的情景实验方法缺乏现实世界的真实性和复杂性。未来研究可以考虑采用实地实验或自然实验等方法,在真实的工作环境中观察和分析算法决策的影响,提高研究结果的外部效度。

第二,本研究中提出的中介机制和边界条件还有待更充分地检验。一方面,除能力和热情两个维度之外,还可能存在其他人们对算法决策评价的特质(例如信任)作为潜在中介变量。另一方面,本研究为操纵任务的主观性特征使用了两种差别较大的管理任务,难免会引入其他干扰因素。未来可以考虑利用同一管理任务场景,通过改变任务的主客观特征对本文的结论进一步加以检验。

第三,本研究只考虑了工作排班和绩效评估两个具体的决策场景。员工在其他场景下是否有类似的公平感知及其背后的作用机制还有待进一步验证。研究者未来可以在更丰富的组织管理任务场景中对现有结果的普适性加以检验。此外,算法决策的应用不仅限于组织管理,还广泛应用于消费、政治、司法、健康等多个领域。未来的研究可以探讨在这些领域中,算法决策如何影响人们的公平感知,以及如何通过优化算法设计来提高其接受度。

第四,本研究暂未探索员工感知公平对员工和组织后续影响。面对算法决策和人类决策的不利结果,员工是否会产生不同的应对方式,如工作投入、组织公民行为、建言、怠工和耗竭?未来的研究可以将算法决策对员工感知公平影响的后效(如员工的应对方式、消极行为和组织绩效等)纳入研究模型中,从而更完整地理解算法决策对组织行为的影响。

附 录

表 A1 前测实验场景描述

工作排班	绩效评估
请设想您是一家连锁快餐店的店长，管理大约 10 名店员。店长的工作职责中有一项任务：在每周末为所有店员下一周的工作排班。快餐店在一天中的客流量波动较大，有些时段比较轻松，有些时段比较辛苦，但不同时段的小时工资是一样的。因此，店长在排班的时候不仅要保证所有店员的工作时长基本一致，还需要把他们公平地分配到轻松时段和繁忙时段	请设想您是一家电商平台的电话客服中心主管，管理大约 10 名客服人员。主管的工作职责中有一项任务：对客服人员每周的工作表现进行考核。客服人员的日常工作主要是接听并处理顾客的咨询和售后电话。主管每周都需要随机抽取每位客服人员的部分电话录音进行检查，根据他们接电话时的语气、语调、态度积极程度对其工作表现给出评价

表 A2 主实验场景描述

任务		人类组	算法组
工作排班	场景描述	请设想您在一家快餐店打工。快餐店一天中的客流量波动很大，有些时段比较轻松，有些时段比较辛苦（例如午餐晚餐时段），但不同时段的小时工资是一样的	
	操纵	这家店由店长每周给员工分配排班。一个月前，老板聘请了一位新店长	这家店原本由店长每周给员工分配排班。一个月前，老板给店里安装了一套基于人工智能算法的自动化排班软件，为员工自动分配每周的排班
	不利结果	过了一段时间，您发现自己每周被安排在辛苦时段的次数要明显多于其他同事	
绩效评估	场景描述	请设想您在一家电商平台的客服中心工作，主要负责接听顾客的咨询和售后电话	
	操纵	客服中心每个月末都会由客服经理对您的工作绩效进行考核。经理会根据您接听电话时的回答是否准确、态度是否亲切、语音是否清晰给出考核结果	客服中心每个月末都会由一套基于人工智能算法的评估系统对您的工作绩效进行考核。人工智能评估系统会根据您接听电话时的回答是否准确、态度是否亲切、语音是否清晰给出考核结果
	不利结果	今年前几个月，您和同事们的考核成绩基本相同。不过，今天您发现自己上个月的考核结果显著低于其他同事。根据您的观察和了解，上个月您和其他人的工作表现在各方面都比较一致	

参考文献

- [1] 蒋路远,曹李梅,秦昕,谭玲,陈晨,2022. 人工智能决策的公平感知[J]. 心理科学进展, 30(5): 1078-1092.
- [2] 李超平,徐世勇,2019. 人工智能(AI)对中国人力资源管理的影响调查报告[EB/OL]. (2019-3-28) [2023-12-1]. <http://www.obhrm.net/data/人工智能对人力资源管理的影响调查报告.pdf>.
- [3] 裴嘉良,刘善仕,钟楚燕,谌一璠,2021. AI算法决策能提高员工的程序公平感知吗?[J]. 外国经济与管理, 43(11): 41-55.
- [4] 宋晓兵,何夏楠,2020. 人工智能定价对消费者价格公平感知的影响[J]. 管理科学, 33(5): 3-16.
- [5] 魏昕,黄鸣鹏,李欣悦,2021. 算法决策、员工公平感与偏差行为:决策有利性的调节作用[J]. 外国经济与管理, 43(11): 56-69.
- [6] AMBROSE M L, SCHMINKE M, 2009. The role of overall justice judgments in organizational justice research: a test of mediation[J]. Journal of Applied Psychology, 94(2): 491-500.
- [7] ARAUJO T, HELBERGER N, KRUIKEMEIER S, DE VREESE C H, 2020. In AI we trust? perceptions about automated decision-making by artificial intelligence[J]. AI & Society, 35(3): 611-623.
- [8] BAI B, DAI H, ZHANG D J, ZHANG F, HU H, 2022. The impacts of algorithmic work assignment on fairness perceptions and productivity: evidence from field experiments[J]. Manufacturing & Service Operations Management, 24(6): 3060-3078.
- [9] BELANCHE D, CASALÓ L V, FLAVIAN C, SCHEPERS J, 2020. Robots or frontline employees? exploring customers' attributions of responsibility and stability after service failure or success[J]. Journal of Service Management, 31(2): 267-289.
- [10] BELANCHE D, CASALÓ L V, SCHEPERS J, FLAVIÁN C, 2021. Examining the effects of robots' physical appearance, warmth, and competence in frontline services: the humanness-value-loyalty model[J]. Psychology & Marketing, 38(12): 2357-2376.
- [11] BINNS R, VAN KLEEK M, VEALE M, LYNGS U, ZHAO J, SHADBOLT N, 2018. 'It's reducing a human being to a percentage'; perceptions of justice in algorithmic decisions[C]. Montreal: ACM Conference on Human Factors in Computing Systems.
- [12] BOGERT E, SCHECTER A, WATSON R T, 2021. Humans rely more on algorithms than social influence as a task becomes more difficult[J]. Scientific Reports, 11(1). DOI: 10.1038/s41598-021-87480-9.
- [13] BROCKNER J, FISHMAN A Y, REB J, GOLDMAN B M, SPIEGEL S, 2007. Procedural fairness, outcome favorability, and judgments of an authority's responsibility[J]. Journal of Applied Psychology, 92(6): 1657-1671.
- [14] BUCKLEY M, 2012. Justice in context: assessing contextualism as an approach to justice[J]. Ethics & Global Politics, 5(2): 71-94.
- [15] CAMPBELL W K, SEDIKIDES C, 1999. Self-threat magnifies the self-serving bias: a meta-analytic integration[J]. Review of General Psychology, 3(1): 23-43.
- [16] CASTELO N, BOS M W, LEHMANN D R, 2019. Task-dependent algorithm aversion[J]. Journal of Marketing Research, 56(5): 809-825.
- [17] CHOI S, 2011. Organizational justice and employee work attitudes: the federal case[J]. American Review of Public Administration, 41(2): 185-204.
- [18] CHOON L K, EMBI M A, 2012. Subjectivity, organizational justice and performance appraisal: understanding the

- concept of subjectivity in leading towards employees' perception of fairness in the performance appraisal[J]. *Procedia-Social and Behavioral Sciences*, 62: 189–193.
- [19] CHOUNG H, SEBERGER J S, DAVID P, 2023. When AI is perceived to be fairer than a human: understanding perceptions of algorithmic decisions in a job application context[J]. *International Journal of Human-Computer Interaction*, 40(22): 7451–7468.
- [20] COJUHARENCO I, MARQUES T, PATIENT D, 2017. Tell me who, and I'll tell you how fair: a model of agent bias in justice reasoning[J]. *Group & Organization Management*, 42(5): 630–656.
- [21] COLQUITT J A, SCOTT B A, RODELL J B, LONG D M, ZAPATA C P, CONLON D E, WESSON M J, 2013. Justice at the millennium, a decade later: a meta-analytic test of social exchange and affect-based perspectives[J]. *Journal of Applied Psychology*, 98(2): 199–236.
- [22] CROPANZANO R, 2012. *Justice in the workplace: from theory to practice*[M]. New York: Psychology Press.
- [23] CUDDY A J C, FISKE S T, GLICK P, 2007. The BIAS map: behaviors from intergroup affect and stereotypes[J]. *Journal of Personality and Social Psychology*, 92(4): 631–648.
- [24] DAUGHERTY P R, WILSON H J, CHOWDHURY R, 2019. Using artificial intelligence to promote diversity[J]. *MIT Sloan Management Review*, 60(2): 10–12.
- [25] DAUS C S, ASHKANASY N M, 2005. The case for the ability-based model of emotional intelligence in organizational behavior[J]. *Journal of Organizational Behavior*, 26(4): 453–466.
- [26] DE CREMER D, 2020. What does building a fair AI really entail? [J]. *Harvard Business Review*. <https://hbr.org/2020/09/what-does-building-a-fair-ai-really-entail>.
- [27] DHIR K, CHHABRA A, 2019. Automated employee evaluation using fuzzy and neural network synergism through IoT assistance[J]. *Personal and Ubiquitous Computing*, 23(1): 43–52.
- [28] DONG M, BONNEFON J F, RAHWAN I, 2024. Toward human-centered AI management: methodological challenges and future directions[J]. *Technovation*, 131(4). DOI: 10.1016/j.technovation.2024.102953.
- [29] ELOVAINIO M, KIVIMAKI M, HELKAMA K, 2001. Organization justice evaluations, job control, and occupational strain[J]. *Journal of Applied Psychology*, 86(3): 418–424.
- [30] FISKE S T, CUDDY A J C, GLICK P, 2007. Universal dimensions of social cognition: warmth and competence [J]. *Trends in Cognitive Sciences*, 11(2): 77–83.
- [31] FISKE S T, CUDDY A J C, GLICK P, XU J, 2002. A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition[J]. *Journal of Personality and Social Psychology*, 82(6): 878–902.
- [32] FOLKES V S, 1988. Recent attribution research in consumer-behavior—a review and new directions[J]. *Journal of Consumer Research*, 14(4): 548–565.
- [33] FUMAGALLI E, REZAEI S, SALOMONS A, 2022. OK computer: worker perceptions of algorithmic recruitment [J]. *Research Policy*, 51(2). DOI: 10.1016/j.respol.2021.104420.
- [34] FURNHAM A, 2001. Self-estimates of intelligence: culture and gender difference in self and other estimates of both general (g) and multiple intelligences[J]. *Personality and Individual Differences*, 31(8): 1381–1405.
- [35] GAL U, JENSEN T B, STEIN M K, 2020. Breaking the vicious cycle of algorithmic management: a virtue ethics approach to people analytics[J]. *Information and Organization*, 30(2). DOI: 10.1016/j.infoandorg.2020.100301.
- [36] GARVEY A M, KIM T, DUHACHEK A, 2023. Bad news? send an AI. good news? send a human[J]. *Journal of Marketing*, 87(1): 10–25.
- [37] GELBRICH K, HAGEL J, ORSINGER C, 2021. Emotional support from a digital assistant in technology-mediated services: effects on customer satisfaction and behavioral persistence[J]. *International Journal of Research in Marketing*, 38(1): 176–193.

- [38] GRANULO A, CAPRIOLI S, GUCHS C, PUNTONI S, 2024 The social cost of algorithmic management[J]. Harvard Business Review. <https://hbr.org/2024/02/the-social-cost-of-algorithmic-management>.
- [39] GRAY H M, GRAY K, WEGNER D M, 2007. Dimensions of mind perception[J]. Science, 315(5812): 619.
- [40] GRAY K, WEGNER D M, 2012. Feeling robots and human zombies: mind perception and the uncanny valley[J]. Cognition, 125(1): 125–130.
- [41] GREENBERG J, 1990. Organizational justice: yesterday, today, and tomorrow[J]. Journal of Management, 16(2): 399–432.
- [42] HAAK T, 2017. Algorithm aversion and HR-Tech[R]. HR Trend Institute. <https://www.hrtrendinstitute.com/blog/2017/02/13/algorithm-aversion-hr-trends-2017-5>.
- [43] HAYES A F, 2013. Introduction to mediation, moderation, and conditional process analysis: a regression-based approach[M]. New York: The Guilford Press.
- [44] HEIDER F, 1958. The psychology of interpersonal relations[M]. New York: Wiley.
- [45] HILLIARD A, GUENOLE N, LEUTNER F K, 2022. Robots are judging me: perceived fairness of algorithmic recruitment tools[J]. Frontiers in Psychology, 13. DOI: 0.3389/fpsyg.2022.940456.
- [46] HÖDDINGHAUS M, SONDERN D, HERTEL G, 2021. The automation of leadership functions: would people trust decision algorithms?[J]. Computers in Human Behavior, 116. DOI: 10.1016/j.chb.2020.106635.
- [47] HOLLENSBE E C, KHAZANCHI S, MASTERS S S, 2008. How do I assess if my supervisor and organization are fair? identifying the rules underlying entity-based justice perceptions[J]. Academy of Management Journal, 51(6): 1099–1116.
- [48] HOSHINO R, SLOBODIN A, BERNOUDY W, 2018. An automated employee timetabling system for small businesses[C]. New Orleans: Thirty-Second AAAI Conference on Artificial Intelligence, 7673–7679.
- [49] HOWARD F M, GAO C A, SANKEY C, 2020. Implementation of an automated scheduling tool improves schedule quality and resident satisfaction[J]. PloS One, 15(8). DOI: 10.1371/journal.pone.0236952.
- [50] JACKSON D N, MESSICK S J, 1957. A note on “ethnocentrism” and acquiescent response sets[J]. Journal of Abnormal and Social Psychology, 54(1): 132–134.
- [51] JAGO A S, CARROLL G R, 2024. Who made this? algorithms and authorship credit[J]. Personality and Social Psychology Bulletin, 50(5): 793–806.
- [52] JUDD C M, JAMES-HAWKINS L, YZERBYT V, KASHIMA Y, 2005. Fundamental dimensions of social judgment: understanding the relations between judgments of competence and warmth[J]. Journal of Personality and Social Psychology, 89(6): 899–913.
- [53] KAIBEL C, KOCH-BAYRAM I, BIEMANN T, MÜHLENBOCK M, 2019. Applicant perceptions of hiring algorithm—uniqueness and discrimination experiences as moderators[J]. Academy of Management Proceedings, 2019(1). DOI: 10.5465/AMBPP.2019.210.
- [54] KANFER R, ACKERMAN P L, 1989. Motivation and cognitive-abilities—an integrative aptitude treatment interaction approach to skill acquisition[J]. Journal of Applied Psychology, 74(4): 657–690.
- [55] KELLEY H H, MICHELA J L, 1980. Attribution theory and research[J]. Annual Review of Psychology, 31: 457–501.
- [56] KELLOGG K C, VALENTINE M A, CHRISTIN A, 2020. Algorithms at work: the new contested terrain of control[J]. Academy of Management Annals, 14(1): 366–410.
- [57] KERVYN N, FISKE S, YZERBYT V, 2015. Forecasting the primary dimension of social perception symbolic and realistic threats together predict warmth in the stereotype content model[J]. Social Psychology, 46(1): 36–45.
- [58] KIM J Y, HEO W, 2022. Artificial intelligence video interviewing for employment: perspectives from applicants, companies, developer and academicians[J]. Information Technology & People, 35(3): 861–878.

- [59] KIM Y, EOM J, YOON T J, CHUNG A J, KIM M, 2021. AI divide versus inclusion: an empirical evidence from an on-demand food delivery platform [C]. Austin: ICIS 2021 Proceedings, 5.
- [60] LAMBERT L S, TEPPER B J, CARR J C, HOLT D T, BARELKA A J, 2012. Forgotten but not gone: an examination of fit between leader consideration and initiating structure needed and received [J]. *Journal of Applied Psychology*, 97(5): 913–930.
- [61] LANGER M, KÖNIG C J, PAPATHANASIOU M, 2019. Highly automated job interviews: acceptance under the influence of stakes [J]. *International Journal of Selection and Assessment*, 27(3): 217–234.
- [62] LANGER M, KÖNIG C J, SANCHEZ D R P, SAMADI S, 2020. Highly automated interviews: applicant reactions and the organizational context [J]. *Journal of Managerial Psychology*, 35(4): 301–314.
- [63] LAU C M, OGER B, 2012. Behavioral effects of fairness in performance measurement and evaluation systems: empirical evidence from France [J]. *Advances in Accounting*, 28(2): 323–332.
- [64] LEE M K, 2018. Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management [J]. *Big Data & Society*, 5(1). DOI: 10.1177/2053951718756684.
- [65] LEE M K, KUSBIT D, METSKY E, DABBISH L A, 2015. Working with machines: the impact of algorithmic and data-driven management on human workers [R]. Seoul: Proceedings of the 33rd Annual CHI Conference on Human Factors in Computing Systems. DOI: 10.1145/2702123.2702548.
- [66] LEI X, RAU P L P, 2021. Effect of robot tutor's feedback valence and attributional style on learners [J]. *International Journal of Social Robotics*, 13(7): 1579–1597.
- [67] LEICHT-DEOBALD U, BUSCH T, SCHANK C, WEIBEL A, 2019. The challenges of algorithm-based HR decision-making for personal integrity [J]. *Journal of Business Ethics*, 160(2): 377–392.
- [68] LERNER M J, 1965. Evaluation of performance as a function of performer's reward and attractiveness [J]. *Journal of Personality and Social Psychology*, 1(4): 355–360.
- [69] LEWIS N, MARC J, 2019. Want to work for L'Oreal? get ready to chat with an AI bot [N]. CNN Business, 4–29. <https://edition.cnn.com/2019/04/29/tech/ai-recruitment-loreal/index.html>.
- [70] LI M, BITTERLY T B, 2024. How perceived lack of benevolence harms trust of artificial intelligence management [J]. *Journal of Applied Psychology*, 109(11): 1794–1816.
- [71] LINDEBAUM D, ASHRAF M M, 2021. The ghost in the machine, or the ghost in organizational theory? a complementary view on the use of machine learning [J]. *Academy of Management review*, 49(2): 445–448.
- [72] LUO X, TONG S, FANG Z, QU Z, 2019. Frontiers: machines vs. humans: the impact of artificial intelligence chatbot disclosure on customer purchases [J]. *Marketing Science*, 38(6): 937–947.
- [73] MCGUIRE J, DE CREMER D, 2023. Algorithms, leadership, and morality: why a mere human effect drives the preference for human over algorithmic leadership [J]. *AI and Ethics (Online)*, 3(2): 601–618.
- [74] MCKEE K R, BAI X, FISKE S T, 2023. Humans perceive warmth and competence in artificial intelligence [J]. *iScience*, 26(9). DOI: 10.1016/j.isci.2023.107603.
- [75] MILKOVICH G T, WIGDOR A K, 1991. Pay for performance: evaluating performance appraisal and merit pay [M]. Washington: The National Academies Press.
- [76] MILLER S M, KEISER L R, 2021. Representative bureaucracy and attitudes toward automated decision making [J]. *Journal of Public Administration Research and Theory*, 31(1): 150–165.
- [77] MOSSHOLDER K W, BENNETT N, KEMERY E R, WESOŁOWSKI M A, 1998. Relationships between bases of power and work reactions: the mediational role of procedural justice [J]. *Journal of Management*, 24(4): 533–552.
- [78] NEWLANDS G, 2021. Algorithmic surveillance in the gig economy: the organization of work through lefebvrian conceived space [J]. *Organization Studies*, 42(5): 719–737.
- [79] NEWMAN D T, FAST N J, HARMON D J, 2020. When eliminating bias isn't fair: algorithmic reductionism and

- procedural justice in human resource decisions[J]. *Organizational Behavior and Human Decision Processes*, 160: 149–167.
- [80] NOBLE S M, FOSTER L L, CRAIG S B, 2021. The procedural and interpersonal justice of automated application and resume screening[J]. *International Journal of Selection and Assessment*, 29(2): 139–153.
- [81] PARK H, AHN D, HOSANAGAR K, LEE J, 2021. Human-AI interaction in human resource management: understanding why employees resist algorithmic evaluation at workplaces and how to mitigate burdens[C]. Yokohama: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.
- [82] PLOYHART R E, RYAN A M, 1997. Toward an explanation of applicant reactions: an examination of organizational justice and attribution frameworks[J]. *Organizational Behavior and Human Decision Processes*, 72(3): 308–335.
- [83] RUSSELL S J, NORVIG P, 2003. Artificial intelligence: a modern approach[M]. 2nd ed. Engelwood Cliff City: Prentice Hall.
- [84] SANDERS N R, WOOD J D, 2020. The humachine: humankind, machines, and the future of enterprise[M]. New York: Routledge.
- [85] SHAVER K G, 1970. Defensive attribution—effects of severity and relevance on responsibility assigned for an accident[J]. *Journal of Personality and Social Psychology*, 14(2): 101–113.
- [86] SKARLICKI D P, FOLGER R, 1997. Retaliation in the workplace: the roles of distributive, procedural, and interactional justice[J]. *Journal of Applied Psychology*, 82(3): 434–443.
- [87] SRINIVASAN R, SARIAL-ABI G, 2021. When algorithms fail: consumers' responses to brand harm crises caused by algorithm errors[J]. *Journal of Marketing*, 85(5): 74–91.
- [88] SUN Z, ZANG G, WANG Z, ZHAO H, 2023. VCAs as partners or servants? the effects of information sensitivity and anthropomorphism roles on privacy concerns[J]. *Technological Forecasting and Social Change*, 192. DOI: 10.1016/j.techfore.2023.122560.
- [89] TAMBE P, CAPPELLI P, YAKUBOVICH V, 2019. Artificial intelligence in human resources management: challenges and a path forward[J]. *California Management Review*, 61(4): 15–42.
- [90] TARAFDAR M, PAGE X, MARABELLI M, 2023. Algorithms as co-workers: human algorithm role interactions in algorithmic work[J]. *Information Systems Journal*, 33(2): 232–267.
- [91] TJOSVOLD D, 1984. Effects of leader warmth and directiveness on subordinate performance on a subsequent task[J]. *Journal of Applied Psychology*, 69(3): 422–427.
- [92] TOMPROU M, LEE M K, 2019. Psychological contracts in algorithmic management[J]. *Academy of Management Proceedings*. DOI: 10.5465/AMBPP.2019.12894.
- [93] TONG S, JIA N, LUO X, FANG Z, 2021. The Janus face of artificial intelligence feedback: deployment versus disclosure effects on employee performance[J]. *Strategic Management Journal*, 42(9): 1600–1631.
- [94] VAN DOORN J, MENDE M, NOBLE S M, HULLAND J, OSTROM A L, GREWAL D, PETERSEN J A, 2017. Domo arigato mr. Roboto: emergence of automated social presence in organizational frontlines and customers' service experiences[J]. *Journal of Service Research*, 20(1): 43–58.
- [95] VON WALTER B, KREMMEL D, JÄGER B, 2022. The impact of lay beliefs about AI on adoption of algorithmic advice[J]. *Marketing Letters*, 33: 143–155.
- [96] VOUBEM L, KRAMER S, SCHÄFFER U, 2016. Fairness perceptions of annual bonus payments: the effects of subjective performance measures and the achievement of bonus targets[J]. *Management Accounting Research*, 30: 32–46.
- [97] WOLOSIN R J, SHERMAN S J, TILL A, 1973. Effects of cooperation and competition on responsibility attribution after success and failure[J]. *Journal of Experimental Social Psychology*, 9(3): 220–235.
- [98] WU X, LIU Q, QU H, WANG J, 2023. The effect of algorithmic management and workers' coping behavior: an ex-

- ploratory qualitative research of Chinese food-delivery platform [J]. *Tourism Management*, 96. DOI: 10.1016/j.tourman.2022.104716.
- [99] XU Y, SHIEH C H, VAN ESCH P, LING I L, 2020. AI customer service: task complexity, problem-solving ability, and usage intention [J]. *Australasian Marketing Journal*, 28(4): 189–199.
- [100] YANG J, MOSSHOLDER K W, PENG T K, 2009. Supervisory procedural justice effects: the mediating roles of cognitive and affective trust [J]. *Leadership Quarterly*, 20(2): 143–154.
- [101] ZHANG L, AMOS C, 2023. Dignity and use of algorithm in performance evaluation [J]. *Behaviour & Information Technology*, 43(2): 401–418.
- [102] ZHANG L, YENCHIA C, 2022. Examining perceptions towards hiring algorithms [J]. *Technology in Society*, 68. DOI: 10.1016/j.techsoc.2021.101848.

The Impact of Algorithmic Decision-Making on Employees' Perceptions of Fairness

Yi Jing

(Postdoctoral Research Center, Industrial and Commercial Bank of China)

Lingyun Qiu* Run Ren

(Guanghua School of Management, Peking University)

Summary: This paper investigates the influence of algorithmic decision-making on employees' perceptions of fairness within the context of organizational management, particularly in light of the rapid advancements in artificial intelligence (AI). Leveraging the Stereotype Content Model (SCM), the study explores differential fairness perceptions between algorithmic and human decision-makers, particularly in adverse outcome scenarios. Findings suggest that algorithms are generally viewed as fairer than human managers, with perceptions influenced significantly by the type of task being evaluated.

As AI technologies continue to permeate various business operations, organizations increasingly deploy algorithms for diverse managerial functions, including human resources management, task allocation, and performance evaluations. This shift raises critical questions about how employees perceive the fairness of decisions made by algorithms as opposed to human managers. Given that fairness perceptions are pivotal to employee satisfaction, organizational commitment, and performance, understanding these dynamics is essential for effective organizational management.

Existing research on algorithmic decision-making offers mixed insights into its impact on perceived fairness. Some studies argue that algorithms, by relying on data and objective models, minimize human biases and enhance fairness. Others, however, highlight potential shortcomings such as neglect of qualitative information and lack of transparency, which can lead to perceived unfairness. Additionally, the literature suggests a gap in understanding the effects of decision outcomes on fairness perceptions, particularly when outcomes are unfavorable, thus forming the basis of this study's inquiry.

This study employs scenario-based experiments to examine how employees perceive fairness when confronted with unfavorable decisions executed by either algorithms or human managers. These experiments are designed to cover a range of task subjectivities, from highly objective, data-driven tasks to those requiring significant human judgment and intuition.

The experimental results reveal that decision-maker type and task nature significantly affect fairness perceptions. In scenarios involving objective tasks, algorithms are perceived as fairer due to

* Corresponding Author: Lingyun Qiu, Guanghua School of Management, Peking University, E-mail: qiu@gsm.pku.edu.cn.

their presumed impartiality and lack of human error. For subjective tasks, algorithms are still viewed more favorably, but this is attributed to human managers being perceived as potentially indifferent or lacking empathy. This dichotomy underscores the complexity of fairness perceptions and suggests that while algorithms may excel in objectivity, they may fall short in areas requiring emotional intelligence.

This research adds depth to the discussion on AI's role in management by delineating how task type and outcome influence fairness perceptions differently under algorithmic versus human decision-making. It offers insights that could help organizations better integrate AI into their management practices, ensuring that fairness perceptions are carefully managed to maintain employee satisfaction and performance.

Future researches could broaden the investigation into other organizational contexts and include longitudinal studies to assess how fairness perceptions evolve with long-term exposure to algorithmic decision-making. Moreover, further studies could explore how increased transparency and employee involvement in algorithm development might enhance trust and fairness perceptions, fostering a more equitable organizational environment.

Keywords: Algorithmic Decision-Making; Fairness Perception; Stereotype Content Models; Task Type

JEL Classification: M10