



不同裁判模式对第三方公众赛事公平性感知的研究： 基于理解错觉的中介和道德理解的调节

邬红丽¹, 唐小飞^{2*}

(1.成都体育学院, 四川 成都 610041; 2.西南财经大学 工商管理学院, 四川 成都 610030)

摘要:随着人工智能技术的发展,以及国际体育赛事中裁判员不公平的裁判事件被媒体不断曝光,越来越多的媒体和公众呼吁采纳人工智能裁判取代传统裁判员。选取2018—2022年国际体育赛事中涉及人工智能裁判模式与传统裁判员模式中有争议的媒体材料,通过三组实验验证有关假设。研究表明:人工智能裁判模式对第三方公众赛事公平性感知的的影响要显著大于传统裁判员模式;理解错觉在不同裁判模式与第三方公众赛事公平性感知之间起部分中介作用;道德理解对主效应之间的关系有显著的调节作用,即道德理解抑制了人工智能裁判模式对第三方公众赛事公平性感知的的影响,而增强了传统裁判员模式对第三方公众赛事公平性感知的的影响。

关键词:人工智能;体育裁判;赛事公平感知;理解错觉

中图分类号:G804.87 **文献标识码:**A

随着人工智能技术的发展,以及国际体育赛事中对裁判结果争议问题的不断曝光。一些媒体和公众认为,为了减少人为干扰和确保赛事结果的公平性,应该加强人工智能裁判在体育赛事中的应用,甚至取代传统裁判员。在“如果奥运会的裁判都是AI,会更公平吗?”的一篇评论中:“国际著名体育专栏作家Melendez针对日本东京奥运会系列不公平事件,专门撰文对采纳人工智能裁判的必要性进行了论述”(百度公众号,2021)。而一篇题为《人工智能丢人大发了:把裁判光头当足球追了一整场》(2021)的报道却对人工智能裁判能否取代传统裁判员提出了质疑。目前,人工智能裁判模式与传统裁判员模式相比,哪种模式能更好地体现公平公正的裁判决策,尚缺少研究论证。

依靠专业人士的技能来决策的良好动机与强调采用人工智能来决策会使算法变得透明的观点,引发了一个悖论(Bonezzi et al., 2022)。一方面,一些媒体和公众希望在体育赛事中更多地借助人工智能来确保裁判结果的公平公正。一些研究认为,人工决策的神秘性必然引发人类对社会公平公正更大的担忧,如法官对被告进行评估时都不会解释他们是如何做出判断的(Cohen, 2015),人力资源总监在做出聘用人员决定时是不会解释他们如何评估候选人的(Klehe et al., 2008),医生在诊断时也不会向患者做出全面清晰的解释(Mangano et al., 2015),体育赛事中裁判也不会解释为什么是这样给运动员评分的(张琪

等,2015)等。因此,人类应该广泛采用人工智能(Wu et al., 2022)。因为人工智能在数据处理、知识更新和远程监控方面相较人类具有巨大优势,而且不容易产生偏见,更不会因为与特定团体或运动员发生利益冲突而失去公平公正。相反,一些专业人士认为,国家级和国际级裁判员是参加过多轮严格的专业技术考评和道德评估才取得相应资格的,我们有理由相信他们的职业操守(盛琦,1984)。另外,也有研究证明,人工智能存在一定的局限性。比如,“裁判创造说”认为,规则总是不稳定的,面对层出不穷的新问题,裁判员只能不断地通过创造规则来补全先前的规则(张乃根,1993),这在人工智能裁判领域很难实现;又如,在花样滑冰和花样游泳等兼具技术与表演性质的项目中(徐建方等,2017),人工智能裁判模式相较传统裁判员模式不具优势。2021年,来自*Nature*子刊*Nature Human Behavior*的一项研究从一个全新的视角为传统裁判员模式的合理性提供了佐证。该研究发现,在医学领域,在人工智能诊疗和

收稿日期:2022-04-28; 修订日期:2022-08-25

基金项目:西南财经大学中央高校业务经费资助项目“管理创新研究团队”项目(JBK1805005);国家社科基金资助项目(22XGL017)。

第一作者简介:邬红丽(1973-),女,教授,硕士,主要研究方向为体育文化和运动训练,E-mail:510847213@qq.com。

*通信作者简介:唐小飞(1974-),男,教授,博士,主要研究方向为创新管理,E-mail:xftang@swufe.edu.cn。

医生诊疗中,病患理解错觉(illusory understanding)的中介作用在放大一方优势的同时也在强化另一方的劣势。也就是说,病患放大优势和强化劣势并非客观事实,而是由理解错觉驱动的,即人类的认知模式正如智能算法中“黑匣子”的感觉(Cadario et al., 2021)。与医疗领域的“黑匣子”理论相同,在体育赛事中,一些媒体和公众将“黑匣子”理论套用到了传统裁判员裁判模式之上,对传统裁判员裁判的抵制是由理解裁判员决策行为的主观困难(认为他们的行为是“黑匣子”的感觉)和对人工智能的理解错觉(认为人工智能更精准、无懈可击)共同驱动的,双重驱动夸大了人工智能的作用,抑制了对传统裁判员的信任。加之一些存在争议的比赛结果,使公众对传统裁判员能否秉持公平的裁决产生质疑,这些耦合效应让第三方公众的错觉意识进一步得到强化。

人工智能裁判模式和传统裁判员模式哪种能更好地体现公平公正的裁判决策,目前仍不明确。按照Woodworth(1939)提出的S-O-R模式,即“外在刺激一个体生理、心理认知一个体心理、行为反应”是人类的一般行为模式。显然,人工智能裁判模式和传统裁判员模式是引发理解错觉这一认知变量的外部刺激因素,而第三方公众赛事公平性感知则是作为心理反应的结果变量。因此,研究不同裁判模式对第三方公众赛事公平性感知的影响差异,以及理解错觉在不同裁判模式感知与第三方公众赛事公平性感知之间的中介效用,有助于我们深入理解三者之间的内在生成机制;而将道德理解作为调节变量引入,探讨道德理解对不同裁判模式感知与第三方公众赛事公平性感知之间关系产生的影响,也有利于我们进一步丰富体育管理的相关理论。

1 研究假设

以往研究指出,社会比较是影响公平感知的重要心理过程(吕飒飒等,2021)。Adams(1965)提出的经典公平理论认为,公平感知是人们在社会交换互动过程中与他人进行比较而形成的公平判断。Fehr等(1999)指出,公平感知是个体在社会情境中在参照他人结果的基础上进行的一种主观判断。根据上述研究,在体育赛事中本文将第三方公众赛事公平性感知定义为:社会公众在参照相关赛事结果的基础上,对内群体成员在赛事情境中遭遇的一种公平性主观判断。Colquitt(2001)认为,公平感知涉及公平、公正、客观和合理与否等可操作性内容。关于人工智能裁判和传统裁判员对第三方公众公平性感知影响的研究,目前在司法领域讨论较多,多数研究认为,无论是人工智能裁判模式还是传统裁判员模式,最终目标都是为了提高事件本身的公平性(杜强强,2016;王林清,2015)。Cadario等(2021)的研究也表明,在没有选择权的条件下,智能算法诊疗方式和医生诊疗方式都能被患者接受;在

有选择权的条件下,由于人类决策的主观知识过程大于客观知识,患者选择医生诊疗方式的意愿要显著大于智能诊疗方式,但研究认为这种决策是一种理解错觉。那么,在体育赛事中,人工智能裁判模式相较于传统裁判员模式是否会对第三方公众赛事公平性感知有更强的影响?据此,本文提出假设:

H1:人工智能裁判模式相较传统裁判员模式带来的第三方公众赛事公平性感知显著更强。

人们对人类决策透明度的感知源于一种信念,即“自省”提供了直接进入人们做出决策的心理过程的途径(Nisbett et al., 1977)。但人们实际上无法接触到他们自己的联想机制(Morewedg et al., 2010)。本质上,人类做出的评估往往就像算法所做的决策一样是一个“黑匣子”。因此,对理解错觉的研究在最近两年引起了学术界的广泛关注。理解错觉是指不依靠实际情况,而仅仅依靠自己的意愿和偏见形成的认知(Schlicker et al., 2021),也可以理解为人们的主观偏见。李少远等(2021)在人与机器之间存在异质性的研究中指出,人的理解错觉存在较强的模糊性和随机性,难以通过准确的模型进行描述。因此,理解错觉往往是一种主观想象,而且人类决策过程的主观知识要大于客观知识,对于理解错觉可以从标准、过程和效果三方面的理解程度来明确其操作性定义(Cadario et al., 2021)。另外,理解错觉是一种认知变量,在研究中一般都将其作为中介变量来使用。比如,Bonezzi等(2022)通过实验法发现理解错觉在招聘模式(人工招聘模式和机器人算法招聘模式)与受聘者偏好之间起部分中介作用,原因是受聘者相信他们更了解自己认可的招聘模式,实际上这也是一种错觉。那么,在体育赛事领域,理解错觉是否同样会在不同裁判模式感知与第三方公众公平性感知之间起部分中介作用?据此,本文提出假设:

H2:理解错觉在不同裁判模式与第三方公众赛事公平性感知之间起中介作用。

道德理解是指人们运用自己理解的道德概念或所吸收的道德知识对某一行为的好与坏、善与恶进行道德评价的过程(郑睦凡等,2013)。也有研究认为,道德理解是指公众对道德主体需要根据“应该做什么”和“为什么应该做”所进行的主观判断(陈少明,2007)。孙开宏等(2013)将体育道德界定为运动员和裁判员在竞赛过程中需遵守公平、守法、仁爱、负责、诚实、正义、牺牲与合作等原则。Krebs(2005)的研究通过开发道德理解的测量量表明确了道德理解的操作性定义,包括“出于利己主义的考量”“出于正义的考量”“出于原则的考量”等内容。在消费者伦理领域,一些学者开始关注人工智能算法错误中的伦理问题,并将道德理解作为调节变量引入研究,证明了对道德理解的有效操控可以对人工智能软件算法错误与选择意愿之间的关系起显著的调节作用(Gal et al., 2017)。姜杨和

马艳红(2022)研究证明,感知伦理型教练员在组织伦理氛围与运动员组织信任之间起调节作用。感知伦理型教练员与本文所述的对裁判模式的道德理解有共通的内涵。综上,在体育赛事领域,当导入“道德理解”后,不同裁判模式与第三方公众赛事公平性感知之间的关系是否也会发生改变?据此,本文提出假设:

H3:道德理解对不同裁判模式与第三方公众赛事公平性感知之间的关系会产生显著的调节效应,即道德理解在人工智能裁判模式与第三方公众赛事公平性感知之间起显著的反向调节作用;反之,道德理解在传统裁判员模式与第三方公众赛事公平性感知之间起显著的正向调节作用。

2 量表开发

通过以下几个步骤来获得测量题项:1)对近年来国外相关方向核心期刊进行全面扫描,根据 Voss 等(2003)的语义差异量表,从10个题项中选取3个题项构建了关于人工智能裁判模式和传统裁判员模式测量量表,例如,“对于我来讲,人工智能裁判模式(或传统裁判员模式)是

有效的/无效的、必要的/不必要的、正规的/不正规的”。根据 Cadario 等(2021)的测量量表构建了关于理解错觉的3个测量题项,根据 Krebs 等(2005)的测量量表构建了关于道德理解的3个测量题项,根据 Colquitt(2001)的测量量表构建了关于第三方公众公平性感知的4个题项(表1);2)请相关专业的博士生将其中的英文测项译成中文,随后,将中文问卷再译成英文,并与原文进行对照;3)为了确保证量表的效度,在尽可能保持原量表的基础上,根据概念框架的需要,结合相关理论概念,以及对来自成都体育学院和西南财经大学体育学院5位体育界资深学者(其中正教授3人、副教授2人)的深度访谈,修正了一些题项的表达,以弥补现有测项的不足;4)以初始题项为基础,采用李克特七级量表形式,收集了21份体育爱好者的有效问卷,针对预调研数据进行探索性因子分析,所有变量的 Cronbach's α 值超过0.70,所有题项间的内部相关系数值均介于0.5~0.8,说明量表有较好的信度。另外,量表的KMO值均大于0.7,并通过 Bartlett 球形检验($P < 0.01$),说明量表有较好的结构效度。

表1 测量量表及来源文献
Table 1 Measurement Scale and Source Literature

变量	测量题项	来源文献
人工智能裁判模式	1. 我感知人工智能裁判模式是有效的; 2. 我感知人工智能裁判模式是必要的; 3. 我感知人工智能裁判模式是正规的。	Voss et al., 2003
传统裁判员模式	1. 我感知传统裁判员模式是有效的; 2. 我感知传统裁判员模式是必要的; 3. 我感知传统裁判员模式是正规的。	Voss et al., 2003
理解错觉	1. 我在很大程度上理解这种裁判模式是根据什么标准来裁判的; 2. 我在很大程度上理解这种裁判模式的裁判过程; 3. 我在很大程度上理解这种裁判模式的裁判效果。	Cadario et al., 2021
道德理解	1. 我认为这种裁判模式会促成“出于利己的考量”; 2. 我认为这种裁判模式会促成“出于正义的考量”; 3. 我认为这种裁判模式会促成“出于原则的考量”。	Krebs, 2005
第三方公众赛事公平性感知	1. 作为第三方公众,我感知裁判是公平的; 2. 作为第三方公众,我感知裁判是公正的; 3. 作为第三方公众,我感知裁判是客观的; 4. 作为第三方公众,我感知裁判是合理的。	Colquitt, 2001

3 实验研究

3.1 实验1:不同裁判模式对第三方公众赛事公平性感知的影响

3.1.1 研究对象

借鉴相关研究(Cohen, 1977; Leenaars et al., 2016),取中等效应量 $f^2=0.35$,显著性水平 $\alpha=0.05$,期望功效 $power=0.8$,通过 G*Power 3.1 软件进行样本量估计,得出实验1总的样本量不少于93个。本文参照陈晨等(2020)的实验研究模式,在实验1中,定向式地在MBA在职学员中招募了120名[其中男性占61.2%,平均年龄为(31.6±7.92)岁,平均受教育年限为(17.3±1.75)年]对体育赛事有一定了解

的学员作为操纵对象参加实验。

3.1.2 实验设计

实验1为单因素两水平被试间设计。采用关键事件法对不同裁判模式进行实验操纵,该方法被广泛用于社会心理学与组织行为学的实验操纵中(陈晨等,2020; Liang et al., 2016)。操纵对象被随机分配到人工智能裁判模式感知组($n=60$)和传统裁判员模式感知组($n=60$)。人工智能裁判模式感知组和传统裁判员模式感知组分别阅读情境刺激材料A和B。在阅读情境刺激材料过程中,被试需要想象(时间控制在5 min内)在未来的体育赛事中,人工智能裁判模式或传统裁判员模式会对各类赛事结果

分别带来什么影响。为了加强操控效果, 被试在阅读完情境刺激材料后还需要尽可能详细地描述自己面对这种情境的体会和感受, 字数不限(时间控制在 20 min 内)。在描述完感受之后, 被试填写了理解错觉和第三方赛事公平感知问卷。最后, 被试报告了性别和年龄两项人口统计学信息。

情境刺激材料 A(根据“虎嗅网”的新闻报道《辅助冬奥冠军训练的, 竟然是 AI》改编): 由小冰公司研发的人工智能裁判与教练系统“观君”, 是人工智能在全球顶级赛事中全程参与训练并辅助取得骄人战绩的首个案例。小冰团队从 2019 年就开始随队训练, 在 3 年多的时间里不断收集运动数据, 首创了“小样本、大任务”的冰雪运动分析模型, 并且最终打造出了首个达到国际级竞技赛事专业要求的人工智能裁判与教练系统“观君”。这与我们过去看到的工程师随队操作计算机运行程序并且展示数据不同, “观君”在实际运行中高度自动化, 能够通过采集到的运动轨迹、出台角度、高远度等多维度的关键数据信息直接进行裁判和打分。“观君”对于空中技巧运动员训练过程中出色的评判能力, 获得了国际雪联的认可, 直接引入到北京冬奥测试赛中, 当上了空中技巧项目的唯一竞赛裁判。在冬奥会预赛执裁过程中, “观君”完成了共计 44 人次的全部裁判工作, 裁判结果的精准度、公平性获得了国际雪联、国家体育总局冬季运动管理中心、领队和教练组的一致认可, 圆满完成了人工智能在全球范围内首次大赛独立执裁任务。

情境刺激材料 B(根据“光明网”的新闻报道《千分之一秒定胜负——访中国首位冬奥会裁判员边维华》改编): 作为一名在体育赛事中深耕了 20 多年, 执裁过 2002 年盐

湖城冬奥会、2006 年都灵冬奥会和 2010 年温哥华冬奥会的资深裁判员, 边维华认为: 中国历来重视对裁判人才的培养, 2008 年北京奥运会后, 全国体育机构就展开针对从业人员的大面积培训, 包括技能、道德和规则等内容。本届北京冬奥会就有数百名国内技术裁判, 也是经过层层选拔出来的。边维华说: “裁判员必须熟练掌握比赛规则, 并且能够灵活运用这是基本的要求。尤其是那些千分之一秒定胜负的比赛, 非常考验裁判员的实时反应和观察力, 因此裁判员不仅要注意观察所有细节, 而且要做到判断准确。当然, 我们作为裁判最基本的要求, 也是最重要的要求, 就是要把这个职业当成一份荣耀和一份责任。”

3.1.3 结果检验

1) 主效应检验。采用 SPSS 25.0 进行单因素方差分析, 分析结果显示: 人工智能裁判模式下 ($M=5.754, SD=0.744$) 和传统裁判模式下 ($M=4.746, SD=1.543$) 的第三方公众赛事公平性感知均值都大于中值 0.4, 且参与者的第三方公众赛事公平性感知在人工智能裁判模式下显著高于在传统裁判模式下 ($F=19.147, P<0.001, \eta^2=0.143$), 假设 H1 得到验证。此外, 性别 ($F=0.123, P>0.01$)、年龄 ($F=0.123, P>0.01$) 及其交互作用 ($F=0.475, P>0.01$) 对第三方公众赛事公平性感知无显著影响。

2) 中介效应检验。采用 Bootstrap 方法检验中介效应 (Efron et al., 1979), 对两种裁判模式作虚拟变量处理, 编码为 0(传统裁判员模式)和 1(人工智能裁判模式)。如表 2 所示, 直接效应和间接效应均显著, 95% 的置信区间不包含 0, 表明理解错觉部分中介了不同裁判模式对第三方公众赛事公平性感知的调节影响, 假设 H2 得到支持。

表 2 理解错觉的 Bootstrap 检验

Table 2 Bootstrap Test of Illusory Understanding

路径	效应量	标准误差	95% 置信区间
人工智能裁判模式→第三方公众赛事公平性感知	0.390	0.183	[0.028, 0.751]
人工智能裁判模式→理解错觉→第三方公众赛事公平性感知	0.641	0.195	[0.298, 1.670]
传统裁判员模式→第三方公众赛事公平性感知	-0.390	0.183	[-0.751, -0.028]
传统裁判员模式→理解错觉→第三方公众赛事公平性感知	-0.641	0.193	[-1.47, -0.291]

3.1.4 讨论

通过验证, H1 得到实验数据支持, 即人工智能裁判模式相对于传统裁判员模式带来的第三方公众赛事公平性感知更强。人工智能技术与体育赛事相结合已然是一种显性趋势, 在充分利用科技红利促进人工智能裁判建设的过程中, 也要充分规制人工智能裁判技术中隐藏的“黑匣子”问题。要限制人工智能裁判的过度化适用, 保障人工智能裁判中传统裁判员的适度介入与参与, 强化基于人机协同裁判的各类措施, 尽可能地消解科技融入体育赛事裁判可能产生的冲突, 以实现人工智能裁判模式在程序和结果上的合理性与合规性。进一步, 实验结果证明了

H2 提出的理解错觉的部分中介作用。同时, 研究结论对 Bonezzi 等 (2022) 的理解错觉研究提供了支持。由此可见, 如果在裁判管理中, 我们对理解错觉带来的较高的人工智能裁判模式评价和较低的传统裁判员模式评价放任不管, 第三方公众对传统裁判员模式的理解错觉有可能不断放大, 并影响到正常的裁判秩序。最后, 本文通过控制性别、年龄等因素增强了实验结果的稳健性。

3.2 实验 2: 道德理解在不同裁判模式与第三方公众赛事公平性感知之间的调节效应

3.2.1 研究对象

基于取中等效应量 $f^2=0.35$, 显著性水平 $\alpha=0.05$, 期望

功效 power=0.8,通过 G*Power3.1 软件进行的功效分析表明,本次研究的最低总样本量为 31 个。73 名[其中男性占比 43.3%,平均年龄(30.74±7.721)岁]MBA 在职班的学员参与了实验 2。

3.2.2 实验设计

73 名被试分为人工智能裁判模式感知组(n=36)和传统裁判员模式感知组(n=37)。

首先,人工智能裁判模式感知组阅读实验 1 中的情境刺激材料 A;传统裁判员模式感知组阅读实验 1 中的情境刺激材料 B。在阅读情境刺激材料过程中,被试需要想象(时间控制在 5 min 内)在未来的体育赛事中,人工智能裁判模式或传统裁判员模式会对各类赛事结果分别带来什么影响。为了加强操控效果,被试在阅读完情境刺激材料后还需要尽可能详细地描述自己面对这种情境的体会和感受,字数不限(时间控制在 20 min 内)。在被试描述完感受后,所有被试在李克特七级量表上对人工智能裁判模式感知、传统裁判员模式感和第三方赛事公平性感知打分。

其次,加入刺激。要求人工智能裁判模式感知组和传统裁判员模式感知组所有被试均阅读道德理解的情境刺激材料 C。在阅读情境刺激材料过程中,被试需要想象(时间控制在 5 min 内)在未来的体育赛事中,道德理解对各类赛事结果分别带来什么影响。为了加强操控效果,被试在阅读完情境刺激材料后还需要尽可能详细地描述自己面对这种情境的体会和感受,字数不限(时间控制在 20 min 内)。在被试描述完感受后,所有被试在李克特七级量表上为第三方公众公平性感知和道德理解等测量题项打分。

最后,被试报告了性别和年龄两项人口统计学信息。

情境刺激材料 C(由《凤凰网》的新闻报道《全北贿赂裁判被扣 9 分罚款 1 亿韩元仍获常规赛冠军》改编):裁判,是体育比赛中负责维持赛场秩序,执行比赛规则的职位或人物。许多国际比赛中的裁判必须从比赛双方的国家(地区)以外的第三国(地区)中选出,以示独立、公正和无利益冲突。比如,足球裁判由于多半穿黑色服装,因此经常称为“黑衣法官”,球类运动中的裁判又叫球证。为了尽可能地保证裁判更加公正、客观,在正式比赛中除了设置了裁判员,还设置了巡边员,可以把巡边员理解为是裁判员的助手。各专业裁判委员会,要求裁判要经常观看专业领域内的比赛,以此达到扩大自己的视野,提高全面观察的能力。要懂得专业领域类比赛的技、战术,只有在临场工作时才能把观察力有效地建立在可能会发生违例或犯规的地方,以此达到提高临场裁判员观察的目的性和判断的预见性。当然,裁判员也有失误的时候,不可能做到百分百完美。《人民日报》曾刊发题为《多点理解,少些苛责》的评论文章。文章指出,裁判员在赛场上是规则的化身,在瞬息万变的赛场之上很难做到完美的判罚,偶有错判、漏判出现,我们应该换位思考,尊重裁判,宽容失误。最近几

年,随着人工智能技术的发展,人工智能裁判开始在各类赛事中得到应用。一些人认为,人工智能裁判模式的应用会不会减少利己主义,更多地出于对原则的考量。

3.2.3 实验结果

采用 Bootstrap 法检验道德理解的调节效应。在进行调节效应检验时,对所涉相关连续变量进行中心化处理,以减少共线性问题。在控制了性别和年龄等其余相关因素后,检验了道德理解在人工智能裁判模式感知与第三方公众体育赛事公平性感知(模型 1),以及传统裁判员模式感知与第三方公众体育赛事公平性感知(模型 2)之间的调节作用。如表 3 所示,对于人工智能裁判模式感知组,模型 1 中人工智能裁判模式感知对第三方公众赛事公平性感知具有显著的正向促进作用($\beta=0.163, P<0.01$),人工智能裁判模式感知与道德理解的交互作用显著($\beta=-0.241, P<0.01$),且置信区间[-0.235, -0.571]不包含 0,表明道德理解在人工智能裁判模式感知与第三方公众赛事公平性感知之间起反向调节作用。模型 2 中传统裁判员模式感知对第三方公众赛事公平性感知具有显著的正向促进作用($\beta=0.247, P<0.01$),传统裁判员模式感知与道德理解的交互作用显著($\beta=0.126, P<0.01$),且置信区[0.075, 0.672]不包含 0,表明道德理解在传统裁判员模式感知与第三方公众赛事公平性感知之间起正向调节作用。H3 得到验证。

表 3 道德理解的调节效应检验

变量	模型 1	模型 2
常数项	4.192**(0.377)	4.083**(0.312)
自变量		
人工智能裁判模式	0.163**(0.052)	
传统裁判员模式		0.247**(0.435)
调节变量		
道德理解	0.225**(0.045)	0.198**(0.047)
交互项		
人工智能裁判模式×道德理解	-0.241**(0.035)	
传统裁判员模式×道德理解		0.126**(0.028)
控制变量		
性别	-0.021(0.138)	0.151(0.139)
年龄	0.110(0.057)	0.113*(0.057)
R ²	0.43	0.406
F	9.342**	7.708**

注:1)括号中数值为标准误差;2)*表示 P<0.05,**表示 P<0.01。

3.2.4 讨论

实验 2 验证了 H3,即道德理解在传统裁判员模式与第三方公众赛事公平性感知之间起正向调节作用。这一研究结果意味着,对于公众夸大人工智能裁判员在体育赛事中的作用,以及贬低传统裁判员在体育赛事中的作用,其深层次原因是公众缺乏对传统裁判员的道德理解,

而可能更多的是一种冲动性行为。有研究认为, 裁判道德本质上是一种具有技术含量的合理规范, 它存在于裁判员及其所属的利益群体之间的相互博弈之中, 裁判员背离道德规范不是因为道德水平低下, 而更多的是因为某些利益关系(李兵, 2009)。虽然, 在各类体育赛事中涉及裁判道德的事件并不常见, 但由于体育比赛参与及观看人数多、涉及面广, 又是媒体关注的敏感性话题, 因此, 容易演变为社会性公共危机事件, 从而使裁判道德陷入扩大化的公众信任危机之中。在不同裁判模式对第三方公众赛事公平性感知概念模型中置入道德理解, 对于化解公众的理解错觉, 使公众对人工智能裁判模式和传统裁判员模式做出更加客观的评价与判断均具有重要的现实意义。当然, 在社会化媒体时代, 如果赛事官方能充分利用微博、公众号和各种新媒体, 对体育赛事中传统裁判员道德理解相关知识和正面事件进行更多和更深入报道, 相信第三方公众对传统裁判员模式带来的公平性感知会显著增强, 同时, 也能有效抑制对人工智能裁判模式对公众公平性感知作用的过度夸大。

3.3 实验3: 第三方公众客观理解的机制解释与排除

与第三方公众理解错觉相对应的第三方公众客观理解是否会削弱第三方公众的赛事公平性感知? 对此, 实验3将主要排除第三方公众客观理解的竞争性解释。

3.3.1 主效应测量

选取120名被试[其中男性占37.5%, 平均年龄为(29.99±7.41)岁], 随机分为人工智能裁判模式感知组($n=60$)和传统裁判员模式感知组($n=60$)。重复实验1主效应

实验和检验的全部过程。采用SPSS 25.0进行单因素方差分析, 分析结果显示: 人工智能裁判模式下($M=5.368, SD=0.721$)和传统裁判模式下($M=4.582, SD=1.565$)的第三方公众赛事公平性感知均值都大于中值0.4, 且第三方公众赛事公平性感知在人工智能裁判模式下显著高于在传统裁判模式下($F=20.476, P<0.001, \eta^2=0.150$), 主效应仍然被验证。

3.3.2 客观理解测量

在完成主效应测量后, 参考Cadario等(2021)的研究提炼了3个关于公众客观理解的判断型测量题项: “裁判只需要根据自己的经验在赛场上报分即可?” “对于普通犯规, 技术犯规和违体犯规的判罚是有不同的制裁原则的?” “对于人工智能来说, 寻找开闭角就是增加摄像头个数?”, 被试须根据设置的题项以“正确”“错误”或“不知道”的方式报告其客观理解。答对一题得1分, 其余不计分。

3.3.3 实验结果

本实验以不同裁判模式为自变量, 第三方公众赛事公平性感知为因变量, 客观理解为中介变量, 采用Bootstrap分析客观理解的中介作用。结果显示(表4), 人工智能裁判模式(或传统裁判员模式)与第三方公众赛事公平性感知之间的直接效应显著, 客观理解在人工智能裁判模式(或传统裁判员模式)与第三方公众赛事公平性感知之间的间接效应均不显著。表明客观理解不能解释观察到的第三方公众赛事公平性感知的差异。因此, 第三方公众的客观理解的替代性解释可以被排除。

表4 客观理解的中介效应(Bootstrap)排除检验

Table 4 Bootstrap Exclusion Test for Objective Understanding

路径	效应量	标准误差	95%CI
人工智能裁判模式→第三方公众赛事公平性感知	1.030	0.225	[0.584, 1.476]
人工智能裁判模式→客观理解→第三方公众赛事公平性感知	-0.020	0.039	[-0.100, 0.066]
传统裁判员模式→第三方公众赛事公平性感知	-1.030	0.225	[-1.476, -0.584]
传统裁判员模式→客观理解→第三方公众赛事公平性感知	-0.020	0.040	[-0.066; 0.102]

3.3.4 讨论

实验3的主要目的是对第三方公众客观理解的机制进行解释与排除, 即研究排除了第三方公众客观理解的中介效应。更强的客观理解可以更好地限制冲动, 引导第三方公众管理好他们的情绪化行为(Ryu et al., 2019), 从而削弱他们由于理解错觉而对传统裁判员的负面评价, 增强他们对人工智能裁判的客观评价。也就是说, 只有排除了客观理解的中介效应, 才能说明在体育赛事中的确是理解错觉在不同裁判模式与第三方公众赛事公平性感知之间起中介作用的。研究结论对Schlicker等(2021)的“理解错觉是指不依靠实际情况, 而仅仅依靠自己的意愿和偏见形成的认知”这一观点提供了支持, 并且有助于赛

事主办方了解如何更好地制定公关策略, 引导第三方公众客观的评价不同裁判模式。另外, 采用Cadario等(2021)提炼的客观理解测量量表, 丰富了该量表的应用场景, 并进一步验证了客观理解作为主观理解的解释与排除机制是有效的。实验还进一步对理解错觉的部分中介效应作了强化论证。

4 结论

研究表明, 人工智能裁判模式相较于传统裁判员模式带来了更强的赛事公平性感知。究其原因, 是理解错觉在不同裁判模式与第三方公众赛事公平性感知之间起部分中介作用的结果。研究还证明, 导入道德理解后, 道德理

解能够有效地调节不同裁判模式下的第三方公众赛事公平性感知。基于上述结论,本研究得到如下启示:

1)随着人工智能技术的兴起,人工智能在体育赛事中的应用会变得越来越广泛,但这并不意味着人工智能裁判模式会取代传统裁判员模式。比如,将激光传感器、3D追踪技术、视觉识别等技术引入体育赛事,不仅有助于传统裁判员提高裁判质量,也能有效地减少争议和不确定性。因此,人工智能裁判模式与传统裁判模式并不是对立的矛盾体,恰恰是一种有益的补充。

2)“黑匣子”问题并不只会发生在传统裁判员模式之下,已有的研究和案例也都证明了人工智能裁判模式也会存在算法失误的问题。一些媒体和公众之所以想象人工智能裁判可以近似完美地代替传统裁判员,首先是由于部分主流媒体和自媒体抓住体育赛事中个别裁判员显失公允的判决事件,通过新闻、评论或视频传播方式夸大对整个裁判员队伍的负面评价,从而使社会公众对传统裁判员模式产生理解错觉和公平危机;其次,由于社会公众对传统裁判员模式的公平危机又反过来促使人们急于寻求新的公平的替代方式,而人工智能裁判模式恰好为满足公众的这种需求提供了契机。因此,要警惕理解错觉在不同裁判模式与第三方公众赛事公平性感知之间产生的影响。

3)偏见或偏好并不是固定不变的,某些特定的情境变量可能改变这些认知。比如,加强对“道德理解”的操纵,社会公众则可能认识到“传统裁判员模式更不容易被人为操控的可能性要显著好于人工智能裁判模式”。而且,对经历各种技能考核和道德评估而获得裁判资格的专业裁判人员来说,会更加珍惜来之不易的荣誉。因此,行业主管部门要加强媒体对裁判员道德理解的宣传报道,引导社会公众和媒体要对裁判员有更多的包容、理解和信任。

参考文献:

百度公众号·莫醒我,2021.如果奥运会的裁判都是AI,会更公平吗?[EB/OL].2021-09-01. <https://baijiahao.baidu.com/s?id=1709669023133612851&wfr=spider&for=pc>.

陈晨,张昕,孙利平等,2020.信任以稀为贵?下属感知被信任如何以及何时导致反生产行为[J].心理学报,52(3):329-344.

陈少明,2007.心安。还是理得?:从《论语》的一则对话解读儒家对道德的理解[J].哲学研究,(10):39-45.

杜强强,2016.合宪性解释在我国法院的实践[J].法学研究,6:107-125.

姜杨,马艳红,2022.组织伦理氛围对运动员体育道德发展的影响研究:基于中介与调节[J].沈阳体育学院学报,(5):96-102.

吕飒飒,孙欣,沈林林,等,2021.群体共同经历影响不公平感知[J].心理学报,(7):773-787.

李兵,2009.试论道德的人学基础[J].思想战线,35(3):48-52,75.

李少远,殷翔,2021.信息物理世界如何实现人机共融协同?[J].上海交通大学学报,(A1):5-6.

人民网,2021.人工智能丢人大发了:把裁判光头当足球追了一整场.[EB/OL]. [2020-12-16]. <https://news.mydrivers.com/1/729/>

729998.htm.

孙开宏,季洵,王坤,2013.青少年运动员体育道德结构及相关特征研究[J].体育科学,3:56-61,80.

盛琦,1984.裁判学发凡[J].体育科学,1:84-89.

王林清,2015.司法裁判中的意义分歧及其消解路径[J].中国法学,1:207-223.

徐建方,李良,2017.我国花样游泳项目发展现状分析与对策建议[J].体育科学,10:51-58.

郑睦凡,赵俊华,2013.权力如何影响道德判断行为:情境卷入的效应[J].心理学报,45(11):1274-1282.

张琪,于涛,2015.裁判自由裁量权的伦理规治[J].体育科学,5:78-87.

张乃根,1993.西方法哲学史纲[M].北京:中国政法大学出版社.

ADAMS J S, 1965. Inequity in social exchange[J]. Adv Exp Social Psychol,2(4):267-299.

BONEZZI A, OSTINELLI M, MELZNER J, 2022. The human black-box: The illusion of understanding human better than algorithmic decision-making[J]. J Exp Psychol Gen, 151(9):2250-2258.

CADARIO R, LONGONI C, MOREWEDGE C K, 2021. Understanding, explaining, and utilizing medical artificial intelligence[J]. Nat Human Behav,5: 1636-1642.

COHEN J, 1977. Statistical Power Analysis for the Behavioral Sciences (Rev. ed.) [M]. New York:Lawrence Erlbaum Associates, Inc:267.

COHEN M, 2015. When judges have reasons not to give reasons: A comparative law approach[J]. Washington Lee Law Rev, 72(2): 483-571.

COLQUITT J A, 2001. On the dimensionality of organizational justice: A construct validation of a measure[J]. J Appl Psychol, 86(3):386-400.

EFRON B, EFRON P A, 1979. Bootstrap methods: Another look at jackknife[J]. Ann Statist,(7):1-26.

FEHR E, SCHMIDT K M, 1999. A theory of fairness, competition, and cooperation[J]. Quart J Econom,114(3):817-868.

FISCHHOFF B, SLOVIC P, LICHTENSTEIN S, 1978. Fault Trees: Sensitivity of Estimated Failure Probabilities to Problem Representation[J]. J Exp Psychol: Human Percept Perform, 4(2): 330.

GAL M S, NIVA E K, 2017. Algorithmic consumers[J]. Harvard J Law Technol,30(2):309-353.

KLEHE U C, KONIG C, RICHTER G M, 2008. Transparency in structured interviews: Consequences for construct and criterion-related validity[J]. Human Perform,21(2):107-137.

KREBS D L, 2005. An evolutionary reconceptualization of Kohlberg's model of moral development[J]. Evolu Perspect Human Dev, (2): 243-274.

LEENAARS C H, ZANT J C, AUSSEMS A, et al., 2016. The leeds food preference questionnaire after mild sleep restriction: A small feasibility study[J]. Physiol Behav, 154(1): 28-33.

LIANG L H, LIAN H, BROWN D, et al., 2016. Why are abusive supervisors abusive? A dual system self-control model[J]. Acad Manag J, 59(4): 1385-1406.

MANGANO M, BENNETT S E, GUNNA J, 2015. Creating a patient-centered radiology practice through the establishment of a diagnostic radiology consultation clinic[J]. Am J Roentgenol,205(1):95-99.

NISBETT R E, WILSON T D, 1977. Telling more than we can know: Verbal reports on mental processes[J]. Psychol Rev, 84:231-259.

PRICE W. N, 2018. Big data and black-box medical algorithms [J]. Science, 28:419-467.

RYU H S, KO K S, 2019. Understanding speculative investment behavior in the Bitcoin context from a dual-systems perspective [J]. *Ind Manag Data Syst*, 119(7): 1431-1456.

SCHLICKER N, LANGER M, ÖTTING S K, et al., 2021. What to expect from opening up 'black boxes'? Comparing perceptions of justice between human and automated agents [J]. *Comput Human Behav*, 122:106837.

VOSS K E, SPANGENBERG E R., GROHMAN B, 2003. Measuring the hedonic and utilitarian dimensions of consumer attitude [J]. *J Mark Res*. 40(3):310-320.

WOODWORTH, 1939. Individual and Group Behavior [J]. *Am J Sociol*, 44(6):823-828.

WU J H, XU F, YU L, et al., 2022. Acceptance of medical treatment regimens provided by AI vs. human [J]. *Appl Sci*, 12:110-126.

How Different Referee Modes Influence the Third-party Public Game Fair Perception: Mediation Based on Illusory Understanding and Moderation Based on Moral Understanding

WU Hongli¹, TANG Xiaofei^{2*}

1. Chengdu Sport University, Chengdu 610041, China;

2. School of Business Administration SWUFE, Chengdu 610030, China

Abstract: With the development of artificial intelligence technology, as well as the media exposure of unfair referees in international sports games, more and more media and the public call for adopting AI referees to replace traditional referees. We selected the media materials in the 2018—2022 international games that involve the controversy between the AI referee and the traditional referee through two groups of experiments to verify the relevant assumptions. The results show that: The impact of AI referee on the third-party public game fair perception is significantly greater than that of traditional referee; the illusory understanding plays a part of mediating role between the different referees and the third-party public game fair perception; moral understanding has a significant moderating effect on the relationship between the main effects, that is, moral understanding inhibits the impact of AI referee on the third-party public game fair perception; on the contrary, moral understanding has enhanced the impact of traditional referee on the third-party public game fair perception.

Keywords: artificial intelligence; sports referee; game fairness perception; illusory understanding

(上接第 54 页)

Scientific Exploration on the Fitness Test of Team Sports —Based on the Analysis of Draft Combine Test of the Four Major American Professional Sports Leagues

WEI Xiaobin¹, CHEN Hui², CHEN Xiaoping^{3,4*}

1. School of Strength and Conditioning Training, Beijing Sport University, Beijing 100084, China;

2. School of sports and health management, Chongqing University of Education, Chongqing 400067, China;

3. China Institute of Sport Science, Beijing 100061, China;

4. Faculty of Physical Education, Ningbo University, Ningbo 315211, China

Abstract: This study analyzed and discussed the scientific and precision fitness test of team sports through the draft combine test of the four major American professional sports leagues. It was found that the draft combine fitness test have good overall predictive effects, and it can provide references for the physical fitness assessments in terms of test content, methods, processes and evaluation criteria. The design and optimization of team sports fitness test in China can learn from the professional league assessment systems, based on the characteristics of sports, the multi-dimensional test can be designed according to the field position and age differences. In the tests, the factors affecting the stability and accuracy of assessment results should be avoided through the standardized arrangement of test contents. In addition, a long-term physical fitness assessment database should be established, so as to help the rapid development of Chinese team sports.

Keywords: fitness test; evaluation criteria; team sports; the four major sports leagues