《心理学报》审稿意见与作者回应

题目：感知不透明性增加职场中的算法厌恶
作者：赵一骏 许丽颖（通讯作者） 喻丰（通讯作者） 金旺龙

第一轮

审稿人 1 意见：

研究通过 4 个实验，探讨了个体对算法的厌恶，感知透明性的中介作用，以及拟人化的调节作用。文章选题重要，切合实际，有一定的理论和实践意义。研究设计合理，写作逻辑清楚。小的修改建议：

意见 1：预估样本量时，power 尽量统一，中介效应应该用预估中介所需的样本量，而不是 t-test。

回应：非常感谢您的建议。根据您的建议，首先我们统一了实验 1、3、4 中使用 G*Power 3.1 预估样本量时的统计检验力 power 为 90%。其次，我们研究了预估中介效应所需样本量的方法，参考 Schönbrodt & Perugini（2013）的研究，并根据我们实验 1 的效应量，以 90%的统计检验力（power）和较窄的稳定性通道（corridor of stability）宽度（w = 0.1）参考其文中的 Table 1，确定中介效应所需最低样本量为 150。

参考文献：

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47(5)*, 609-612.

意见 2：在文章中，提到人们的算法厌恶都是从员工的角度来阐释的，比如不公平的感知，而在研究使用意愿是问被试作为企业负责人，这里的不一致需要解释、说明。

回应：非常感谢您的建议。我们在“6.1 理论贡献和实践启示”的部分对您所指出的问题进行了讨论和解释。简单来说，一方面，我们认为从不同职场角色的视角考察人们对算法的态度是具有创新性和必要性的，因为其能更加完整、全面地反应人们对于算法进入职场决策的真实态度。另一方面，视角的差异的确也可以解释我们研究结果中出现的不一致现象。

具体添加内容如下：本研究通过态度的不同维度以及不同的职场视角选取因变量测定了人们对职场中算法使用的厌恶倾向.....此外，本研究在因变量选取上考虑到职场中员工与管理者的不同视角，其中容许性和喜爱程度是从员工的角度来看待算法，而利用意愿则是让被试想象自己作为企业负责人对算法的接受性反应。这一视角上的差异或许也可以解释为什么在实验 1 和实验 4 中利用意愿并没有达到显著性水平。或许原因就在于我们选取的被试比较年轻（一部分为学生被试），其并没有作为企业负责人的生活经验和体会，或许不能代表这一身份的选择倾向性。

意见 3：在研究的 overview 部分，尽可能不要用验证，这样确定的词汇。

回应：非常感谢您的建议。我们非常赞成对于行为用词的严谨态度。因此，我们对文章“1.4 研究概览”部分做了如下调整。将原本“验证”这类具有确定性含义的词汇更改为“探索”、“探求”、“试图发现”等词汇，并添加“将会”、“潜在”等词汇辅以说明，充分体现在提出假设阶段我们并不能完全确定效应是否成立的情况。

具体更改如下：将“这一效应受到感知透明性的中介和拟人化的调节。本研究采用递进的 4 个情境实验来验证此假设”改为“这一效应将会受到感知透明性的中介和拟人化的调节。本研究采用递进的 4 个情境实验来探索此假设的有效性”；将“实验 1 验证了主要假设，即人们表现出对职场中算法 HR 决策的厌恶反应。实验 2 探究其中的内在心理机制，验证感知透明性在决策主体影响决策态度中的中介作用。实验 3 通过操纵算法决策的透明度，进一步检验感知透明性是否是导致人们对算法 HR 决策产生厌恶的前因。实验 4 探索职场中人力决策主体对人们的决策态度影响可能的边界条件，验证拟人化在决策主体影响决策态度中的调节效应。”改为“实验 1 探索主要假设，即人们表现出对职场中算法 HR 决策的厌恶反应。实验 2 探究其中的潜在的心理机制，试图发现感知透明性在决策主体影响决策态度中的中介作用。实验 3 通过操纵算法决策的透明度，进一步考察感知透明性是否是导致人们对算法 HR 决策产生厌恶的前因。实验 4 探索职场中人力决策主体对人们的决策态度影响可能的边界条件，探求拟人化在决策主体影响决策态度中的调节效应。”

意见 4：可以尝试把可容许性、喜欢、意愿作为一个整体的态度指标来进行分析。
回应：非常感谢您的建议。我们审慎地考虑了您的建议，但由于我们也没有明确理论能够将其直接合并为态度指标进行计算分析。于是，我们便对 4 个实验的数据分别进行了多元方差分析以回答您的问题。
具体结果如下：“以决策主体（人类 vs. 算法）为自变量，反映人们对算法态度的可容许性、喜爱程度、利用意愿三个指标得分为因变量进行多元方差分析（MANOVA）。结果表明，决策主体的主效应显著，Wilks’λ = 0.896，$F(3,299) = 11.615$，$p < 0.001$，$\eta_p^2 = 0.104$。”、“以决策主体（人类 vs. 算法）为自变量，性别和年龄为协变量，反映人们对算法态度的可容许性、喜爱程度、利用意愿三个指标得分为因变量进行多元方差分析（MANOVA）。结果表明，决策主体的主效应显著，Wilks’λ = 0.925，$F(3,175) = 4.730$，$p = 0.003$，$\eta_p^2 = 0.075$。”、“以透明性为自变量，性别和年龄为协变量，反映人们对算法态度的可容许性、喜爱程度、利用意愿三个指标得分为因变量进行多元方差分析（MANOVA）。结果表明，透明性的主效应显著，Wilks’λ = 0.907，$F(3,176) = 6.035$，$p = 0.001$，$\eta_p^2 = 0.093$。”、“以不同拟人化程度的决策主体（人类 vs. 拟人化算法 vs. 非拟人化算法）为自变量，对算法熟悉程度、对算法了解程度为协变量，反映人们对算法态度的可容许性、喜爱程度、利用意愿三个指标得分为因变量进行多元方差分析（MANOVA）。结果表明，决策主体的主效应显著，Wilks’λ = 0.965，$F(6,1084) = 3.251$，$p = 0.004$，$\eta_p^2 = 0.018$。”。总而言之，经过 MANOVA，我们发现自变量的主效应已经达到了显著性水平。

意见 5：没有明确的理由可以不用控制性别和年龄。
回应：非常感谢您的建议。由于关于年龄、性别对算法接受度的研究并没有一致的结论，并且我们研究的样本量年龄普遍年轻化，所以的确没有很大必要控制年龄和性别。于是，我们接受了您的建议，删除了每个实验中以年龄和自变量为协变量的方差分析。

意见 6：不需要做 sobel test，重复。
回应：非常感谢您的建议。我们也认为对于中介分析而言，进行 bootstrap 的检验和逐步回归的图示已经足够说明其效应，没必要再进行 sobel test。我们已经在“3.2.2 感知透明性的中介效应”及图示部分中删除 sobel test 的相关内容，只保留了 bootstrap 方法。

意见 7：尽快 direct effect 不显著了，除非有足够的理论和实证证据，也不能轻易下完全中介的结论，这很可能会收到样本的影响。

回应：非常感谢您的建议。根据您的修改意见，我们已经在“3.2.2 感知透明性的中介效应”中删除了由直接效应不再显著得出完全中介效应的有关表述。

审稿人 2 意见：

本研究考察了职场中的算法厌恶，并检验了透明性的中介作用和拟人化的调节作用。研究角度新颖独特，实验设计规范，数据分析相对合理，我认为具有潜在的发表可能性。我有一些建议与作者分享：

意见 1：在样本量分析中，不清楚效应量和检验力的取值是如何确定的。比如实验 2 说参考实验 1 的效应量，那应当选择最小的效应量。

回应：非常感谢您的建议。我们根据您们的建议，做出了统一的修改，即统一每一个实验中预估的检验力（90%）和效应量（均取中等效应量，$d = 0.5$，$f = 0.25$）。并对实验 2 中介分析的样本量参考 Schönbrodt & Perugini（2013）对 Monte Carlo 模拟法的研究，选取较严格的统计检验力 power = 90%，effect size = 0.6（取实验 1 中主要因变量“可容许性”的结果），以及较窄的稳定性通道（corridor of stability）宽度 $w = 0.1$，确定了最小需要样本量为 150。

参考文献：

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47(5),* 609-612.

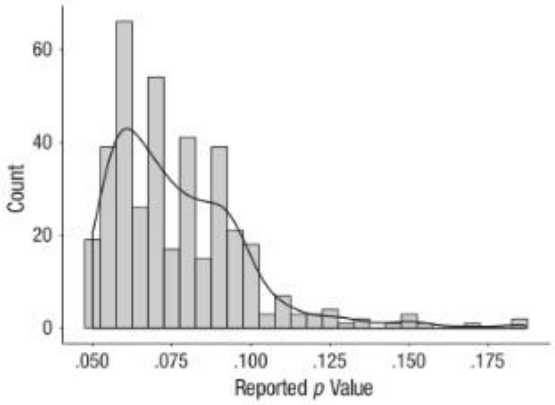
意见 2：建议不要使用边缘显著的说法，没有达到预先设定的显著性水平即为不显著。

回应：非常感谢您的建议。我们对文中提及“边缘显著”的部分进行了如下形式的修改，即将“边缘显著”改换为“不显著，但呈现显著趋势”、“不显著但接近显著”的说法。这一方面能够指出，数据的确没有达到预先设定的显著性水平；另一方面又明确指出，这些变量之间存在接近（approaching）显著的可能性，以便后续研究者更清楚的认识变量之间的关系。Rosnow & Rosenthal（1989）曾在反思心理学家对 p 值的使用时评论道：“Surely, God loves the .06 nearly as much as the .05（诚然，上帝对 0.06 的喜欢几乎和对 0.05 是一样多的）”。Prischet et al.,（2016）的研究指出：“边缘显著的概念几乎已经进入了该领域的每一本实证研究”，并且发现 459 篇文章中标记为边缘显著的 p 值大小有 92.6%处在 0.05-0.10 之间（见下图）。研究者们发现，相比于认知心理学和发展心理学，社会心理学领域报告边缘显著的文章比例更高，且在近 40 年的发展过程中不断升高。这些证据都在说明研究者们越来越愿意通过描述边缘显著效应来支持自己的假设。考虑到贸然报告边缘显著性所带来的种种问题以及科学的严谨性，我们完全采纳您的建议，取消了边缘显著的说法，但将其转换为“不显著+有显著趋势/接近显著”的表达方式。

参考文献：

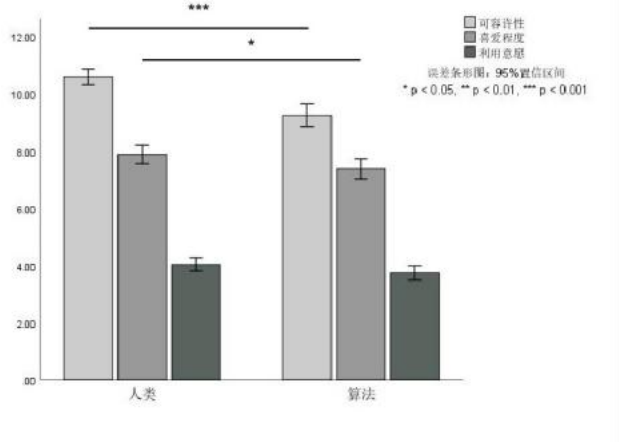
Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44,* 1276–1284.

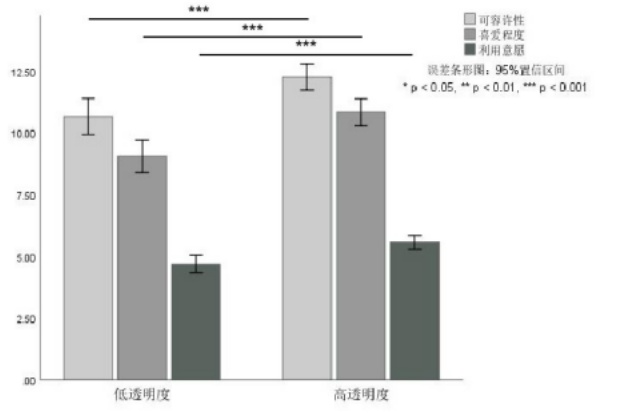
Pritschet, L., Powell, D., & Horne, Z. (2016). Marginally significant effects as evidence for hypotheses: Changing attitudes over for decades. *Psychological Science, 27(7),* 1036-1042.



意见 3：一些结果可考虑用图的形式呈现。

回应：非常感谢您的建议。我们对实验 1 和实验 3 的主要结果进行了图示表达。以下两张柱状图分别是对实验 1 和实验 3 主要结果的可视化展示。





意见 4：很多结果都采用了 2 种分析方法，比如先做 t 检验，然后又把性别年龄作为协变量做方差分析，其实光做后者已经足够，又如中介分析既做了 Bootstrap 又做了 Sobel，其实前者就够了。

回应：非常感谢您的建议。我们删除了协方差分析和 sobel 检验，只保留了 t 检验，和 bootstrap 的中介分析方法。赘余部分已经删除。

意见 5：讨论略显混乱，建议使用小标题。同时，对于未来研究可以从哪几方面展开，可以再阐述得更深入具体一些。

回应：非常感谢您的建议。我们对文章讨论部分进行了重新整理，并划分出两个小标题，即“6.1 理论贡献与实践启示”和“6.2 研究局限与未来展望”，前者主要探讨本文的主要结论有哪些理论和实践价值，并且解释了一些有意思、不一致的发现，而后者则主要描述了本文目前存在的不足，并基于此为后来研究者的工作指明一些可能的方向。

在“6.1 理论贡献与实践启示”中，我们首先回顾了 4 个联系实验的主要结论，并通过强调我们的研究内容和方法涵盖了人力资源管理领域的主要决策场景和具有代表性的被试，以此说明我们发现的研究主题是具有较好稳健性的。其次，我们分别从主效应、中介效应、调节效应的角度讨论了本研究结论与先前研究的一致性，说明了本研究对相关领域的扩展性贡献。另外，我们着重讨论了本研究在因变量选取上的创新意义。我们通过两视角（职场中员工-负责人）和三维度（认知-情感-行为意图）的因变量衡量了人们对算法使用的态度倾向，并且解释了实验中的一部分不一致和阴性结果。

在“6.2 研究巨献与未来展望”中，首先我们指出即便本研究在因变量选取上有所创新，但其仍然难以覆盖人对算法复杂多维的态度，对于后续研究而言，在不同场景下不同的因变量考虑也是有必要、有价值的。其次，我们讨论了本研究中所选取的拟人化方式及其局限性。由于拟人化概念的复杂，不同层次的拟人化可能会引起人们截然不同的反应。然而，本研究仅仅证明了浅层拟人化的有效作用，但没有继续深入探讨更复杂形式拟人化所引起的可能效应。最后，我们还指出，对于职场中人们的算法厌恶可能还存在其他的解释机制和边界条件有待后续研究者深入考察。

意见 6：英文摘要写作不够专业，建议找人润色。

回应：非常感谢您的建议。我们重新根据《心理学报英文摘要写作注意事项》的要求对原由英文摘要进行了润色，精简了表达，目前字数为词符合学报要求。目前英文摘要如下，请审稿专家再次审阅：

In the dynamic arena of modern technological development, algorithms stand out as pivotal tools, seamlessly intertwining with various facets of our digital experiences and societal management. Algorithms as alternatives and enhancements to human decision-making have become ubiquitously adopted in the workplace. Despite algorithms having numerous advantages, such as expansive data storage, expeditious decision-making, objectivity, fairness, and resilience against extraneous disruptive factors, persistent resistance towards algorithmic decision-making is common in existing research. Particularly in the area of human resources, there is often a desire for transparency and openness in decision-making, which is a flaw for algorithms. Thus, the present study aims to explore laypeople's attitudes towards algorithmic management in human resource management and seek the deep mechanism and potential boundary condition.

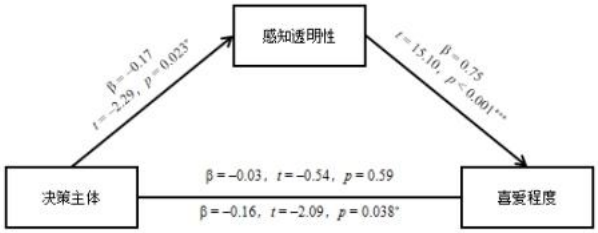
To verify the research hypotheses, four experiments (N=1211) were conducted referring to human resource scenarios in the workplace - recruitment and hiring, allocation of year-end bonuses, resume screening, and performance assessment. Experiment 1 employed a single-factor two-level between-subjects design, wherein 303 participants were randomly assigned to disparate conditions (decision agents: human or algorithm) and tested their permissibility, liking, and willingness to decision-making agents. Experiment 2 was similar to Experiment 1, the only difference was an additional measurement of the mediating role of perceived transparency.

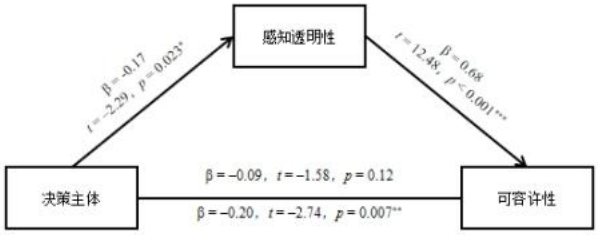
Experiment 3 aimed to reveal the causal linkage between the mediator and dependent variables by manipulating algorithmic decision transparency. In experiment 4, a single-factor three-level (non-anthropomorphic algorithm; anthropomorphic algorithm; human manager) between-subjects design was utilized to assess the moderating effect of anthropomorphism, specifically focusing on manipulating algorithmic anthropomorphism and measuring response propensities.

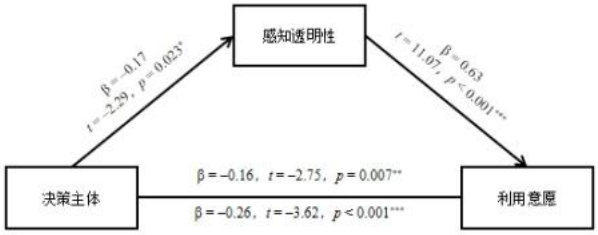
Conforming to predictions, the research revealed a lower degree of permissibility, liking, and to utilize algorithmic decisions compared to those made by human managers (Experiment 1-4). It was robustness that was substantiated across diverse contexts and samples. Furthermore, the study revealed perceived transparency as a mediating variable influencing the impact of decision agents (human vs. algorithm) on three-dimensional (cognitive, affective and behavioral) dependent variables. Specifically speaking, algorithmic decisions, especially concerning human resource matters, were perceived as less transparent and more enigmatic, resulting in their rejection and disapproval (Experiment 2). Algorithms with higher transparency (process and basis for decision-making by open algorithms) would cause a high level of permissibility, liking, and willingness to utilize (Experiment 3). Remarkably, anthropomorphic algorithms—attributed with human-like names and employing human-like linguistic styles—could bring more positive responses, with participants manifesting significantly elevated levels of permissibility, liking, and willingness to utilize the anthropomorphic algorithm (Experiment 4).

These findings bridge the gap in algorithm aversion and decision transparency from the social-psychological perspective. Firstly, this study investigates the different reactions to human-algorithm decisions from a three-dimensional perspective including cognition, affect, and behavioral intention, which could assist in comprehending algorithm aversion. Secondly, the research discovers that perceived transparency is pivotal for algorithm aversion and offers theoretical evidence for the development of Explainable Artificial Intelligence (XAI). Ultimately, showcasing that anthropomorphism can potentially modulate algorithm aversion not only fortifies understanding of psychological triggers but also potentially hints at a pragmatic pathway toward intelligent management.

意见 7：图 1 不规范，建议参考期刊论文中的中介图的画法。
回应：非常感谢您的建议。我们注意到原本中介效应图中出现的圆角方框其应该代表潜变量，而直角方框应为显变量（即直接可以测量的变量）。因此，我们重新绘制了中介效应图示，并删除了其中有关 sobel test 的相关数据，并调整了三个中介效应图示的排列方式，由原本三个中介效应图呈品字形排布转变为竖列排布。如下：







第二轮

审稿人 2 意见：我认为修改稿没有很好地回应我的部分问题。

最主要的问题是，对于作者关于使用“边缘显著”的辩解我不能认同。作者所提到的 2016 年以前有很多学者使用这一说法不能证明使用边缘显著的合理性和正当性。事实上，近年来不少期刊在投稿指南中明确指出不能使用边缘显著的说法。我认为这会导致数据分析更大的灵活性，因为作者可以根据结果是否符合预期来选择“不显著”或“边缘显著”。修改稿中“接近显著”、“有显著趋势”的说法亦不可接受。在修改稿“5.2 结果”一节，甚至将 p=0.055 视为显著。

回应：

非常感谢您的建议。十分抱歉我们上一版的修改稿没有让您满意，这一次我们进行了更加详尽的修订，希望能够达到您的要求。我们非常赞同您对于拒绝使用“边缘显著”这一说法的担忧，这的确会导致研究者在分析数据时有更大的灵活空间进行有利于假设方向的解释。而且我们认为这一条退稿理由是可以修改的，并非是因无法修改而导致的无法接收的问题。

为使表述更加严谨，我们在修改稿中改变了说法。对于这一做法，我们有如下四点理由：首先，我们认为在给定样本量标准上，严苛地执行显著性标准是必要的，但在没有达到显著性时检验程度依然有大小之分，*F* 检验或 *t* 检验并非是一个全或无的值，而是一个连续变量。其次，这种说法在社会心理学领域英文顶刊中多次出现（例如在近期 Journal of Personality and Social Psychology 中，如：Amiot et al., 2020; Essien et al., 2021; Georgeac & Rattan, 2023; Howe et al., 2021; Jiang et al., 2020; 在近期 Journal of Experimental Psychology: General 中，

如：Karmarkar & Kupor, 2023; Righetti et al., 2020; 在近期 Personality & Social Psychology Bulletin 中，如：Huang et al., 2023; Prinzing et al., 2023）；再次，《心理学报》近年发表的论文也有将接近显著性标准的 *p* 值（0.05 < *p* < 0.1）报告为边缘显著或接近显著性水平的做法（例如陈颖 等, 2019; 程亚华 等, 2024; 衡书鹏 等, 2017; 曾欣然 等, 2019; 张清芳 等, 2021）；最后，我们认为将处于 0.05-0.1 之间的显著性水平 *p* 值表达为“不显著但接近显著标准”且同时报告效应量，既可以指出结果的显著性水平确实未达到预先设定的标准，又可以指出变量间潜在可能的关系，以及还可以说明效应如何。

基于此，我们对文章进行了如下两点修改：第一，在实验 1 中，对以利用意愿为因变量的 *t* 检验结果进行了修订，将 0.081 的 *p* 值标定为不显著，但同时强调其 *p* 值小于 0.1，产生较小的效应量。第二，在实验 4 中我们做了三点改动，首先我们将喜爱程度在拟人化与非拟人化算法上的差异表述为接近统计学意义上的显著性水平但不显著，同时也产生较小的效应量；其次我们将以可容许性为因变量的协方差分析结果报告为不显著但接近接近显著标准；最后，我们指出了算法熟悉、了解程度对因变量的积极影响，并在总讨论部分对其进行更深刻的延伸。

以上所参考文献：

① 来源于 JPSP：

Amiot, C. E., Sukhanova, K., & Bastian, B. (2020). Social identification with animals: Unpacking our psychological connection with other animals. *Journal of Personality and Social Psychology, 118*(5), 991–1017.

Essien, I., Calanchini, J., & Degner, J. (2021). Moderators of intergroup evaluation in disadvantaged groups: A comprehensive test of predictions from system justification theory. *Journal of Personality and Social Psychology, 120*(5), 1204–1230.

Georgeac, O. A. M., & Rattan, A. (2023). The business case for diversity backfires: Detrimental effects of organizations' instrumental diversity rthetoric for underrepresented group members' sense of belonging. *Journal of Personality and Social Psychology, 124*(1), 69–108.

Howe, L. C., Carr, P. B., & Walton, G. M. (2021). Normative appeals motivate people to contribute to collective action problems more when they invite people to work together toward a common goal. *Journal of Personality and Social Psychology, 121*(2), 215–238.

Jiang, T., Chen, Z., & Sedikides, C. (2020). Self-concept clarity lays the foundation for self-continuity: The restorative function of autobiographical memory. *Journal of Personality and Social Psychology, 119*(4), 945–959.

② 来源于 JEPG：

Karmarkar, U. R., & Kupor, D. (2023). The unlikelihood effect: When knowing more creates the perception of less. *Journal of Experimental Psychology: General, 152*(3), 906–920.

Righetti, F., Schneider, I., Ferrier, D., Spiridonova, T., Xiang, R., & Impett, E. A. (2020). The bittersweet taste of sacrifice: Consequences for ambivalence and mixed reactions. *Journal of Experimental Psychology: General, 149*(10), 1950–1968.

③ 来源于 PSPB：

Huang, N., Zuo, S., Wang, F., Li, Y., Cai, P., & Wang, S. (2023). New technology evokes old memories: Frequent smartphone use increases feeling of nostalgia. *Personality & Social Psychology Bulletin, 49*(1), 138–151.

Prinzing, M., Van Cappellen, P., & Fredrickson, B. L. (2023). More than a momentary blip in the universe? Investigating the link between religiousness and perceived meaning in life. *Personality and Social Psychology Bulletin, 49*(2), 180–196.

④ 来源于心理学报：

陈颖, 李锋盈, & 李伟健. (2019). 个体关于加工流畅性的信念对字体大小效应的影响. *心理学报, 51*(2), 154–162.

程亚华, 沈岚岚, 李宜逊, 伍新春, 李虹, 王铁群, & 程芳. (2024). 家庭阅读环境对学龄儿童汉字识别、口语词汇知识与阅读理解的影响：一个发展级联模型. *心理学报, 56*(1), 83–92.

衡书鹏, 周宗奎, 牛更枫, & 刘庆奇. (2017). 虚拟化身对攻击性的启动效应：游戏暴力性、玩家性别的影响. *心理学报, 49*(11), 1460–1472.

曾欣然, 汪玥, 丁俊浩, & 周晖. (2019). 班级欺凌规范与欺凌行为：群体害怕与同辈压力的中介作用. *心理学报, 51*(8), 935–944.

张清芳, 钱宗愉, & 朱雪冰. (2021). 汉语口语词汇产生中的多重音韵激活：单词翻译任务中的 ERP 研究. *心理学报, 53*(1), 1–14.

具体修改内容如下，重点部分以红色标出：

（1）第一处相关修改在实验 1 的结果部分：

此外，以利用意愿作为因变量，发现人类组被试对人类 HR 的利用意愿（$M = 4.04$，$SD = 1.42$）高于算法组被试对算法 HR 的利用意愿（$M = 3.74$，$SD = 1.50$），$t(301) = 1.75$，$p = 0.081$，Cohen's $d = 0.21$，差异不显著，但 $p < 0.1$，产生了较小的效应量（Cohen, 1969）。以上结果见图 1 所示。

（2）第二处相关修改在实验 4 的结果部分：

以可容许性作为因变量进行单因素方差分析发现，决策主体的主效应显著，$F(2,546) = 3.15$，$p = 0.044$，$\eta_p^2 = 0.11$；以喜爱程度作为因变量进行单因素方差分析发现，决策主体的主效应不显著，$F(2,546) = 2.39$，$p = 0.093$，$\eta_p^2 = 0.09$；以利用意愿作为因变量进行单因素方差分析发现，决策主体的主效应不显著，$F(2,546) = 0.40$，$p = 0.668$。

计划对比（planned contrast）分析表明，当以可容许性作为因变量时，人类组得分（$M = 10.54$，$SD = 2.81$，$p = 0.021$，Cohen's $d = 0.25$）显著高于非拟人化组算法得分（$M = 9.81$，$SD = 3.11$），而拟人化算法组得分（$M = 10.42$，$SD = 2.94$）在数值上高于非拟人化算法组得分（$M = 9.81$，$SD = 3.11$，$p = 0.055$，Cohen's $d = 0.20$），差异不显著但表现出接近统计学意义的显著性水平，并产生较小的效应量（Cohen, 1969），另外，人类组与拟人化算法组之间并无差异，$p = 0.705$；当以喜爱程度作为因变量时，人类组平均得分（$M = 9.13$，$SD = 3.16$）虽高于非拟人化算法组平均得分（$M = 8.68$，$SD = 3.43$），但未达到统计显著性标准，$p = 0.192$，没有表现出具有统计学意义的算法厌恶。而拟人化算法组得分（$M = 9.40$，$SD = 2.94$，$p = 0.032$，Cohen's $d = 0.23$）显著高于非拟人化算法组得分（$M = 8.68$，$SD = 3.43$），表明算法拟人化对于提高人们对算法决策的喜爱程度是有作用的。（上述结果如下图 4 所示）

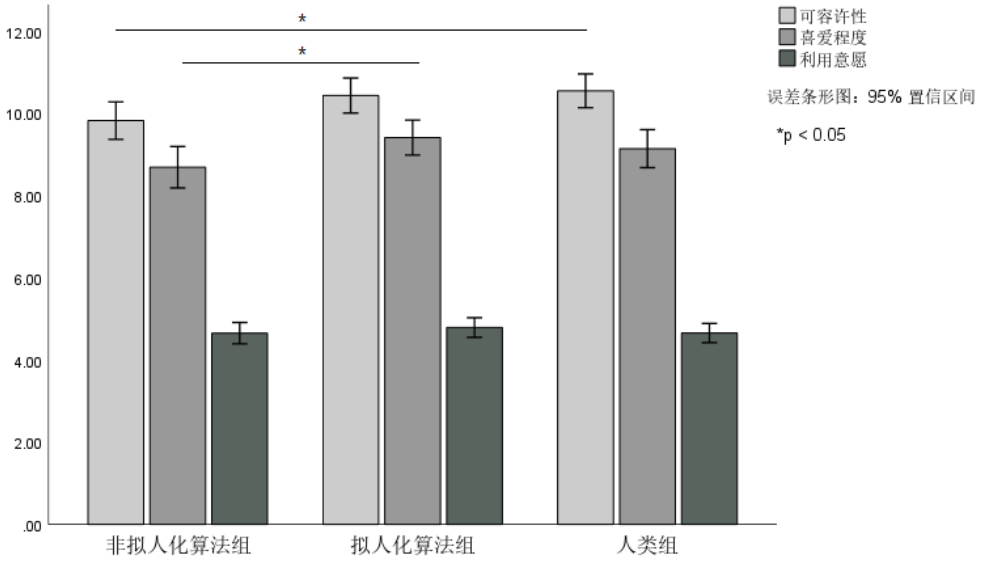


图 4 不同决策主体组的结果比较

以不同拟人化程度的决策主体（人类 vs. 拟人化算法 vs. 非拟人化算法）为自变量，以对算法熟悉程度、对算法了解程度为协变量，以反映人们对算法态度的可容许性、喜爱程度、利用意愿三个指标的得分为因变量进行多元方差分析（MANOVA）。结果表明，决策主体的主效应显著，Wilks'λ = 0.965，$F(6,1084) = 3.251$，$p = 0.004$，$\eta_p^2 = 0.018$。

当将被试报告的对算法的熟悉、了解程度作为协变量，以组别为自变量，当以可容许性为因变量时，进行方差分析，其结果显示，对算法熟悉程度：$F(1,544) = 4.16$，$p = 0.042$，$\eta_p^2 = 0.008$；对算法了解程度：$F(1,544) = 5.64$，$p = 0.018$，$\eta_p^2 = 0.010$，对算法的熟悉、了解程度的效应显著；组别的效应差异不显著但十分接近显著性标准，并产生较小的效应量，$F(2,544) = 2.98$，$p = 0.052$，$\eta_p^2 = 0.011$。保持自变量、协变量不变，当以喜爱程度为因变量时，结果显示，对算法熟悉程度：$F(1,544) = 9.45$，$p = 0.002$，$\eta_p^2 = 0.017$；对算法了解程度：$F(1,544) = 9.67$，$p = 0.002$，$\eta_p^2 = 0.017$；组别的效应同样接近显著性标准，并产生较小的效应量，$F(2,544) = 2.62$，$p = 0.074$，$\eta_p^2 = 0.01$。由此可知，关于算法的先验性的熟悉、了解程度对于人们对待算法管理的态度有极大的影响，这与前人研究保持一致（Ireland, 2020; Komatsu, 2016）。这说明当人们对算法越加熟悉和了解，就越会接纳算法在职场中的各种决策应用，越赋予算法决策以合法地位。换言之，这说明提升民众的算法意识和算法素养同样有助于改善人们的厌恶倾向。

增补的参考文献：

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Ireland, L. (2020). Who errs? Algorithm aversion, the source of judicial error, and public support for self-help behaviors. *Journal of Crime and Justice, 43*(2), 174–192.

Komatsu, T. (2016). Japanese students apply same moral norms to humans and robot agents: Considering a moral HRI in terms of different cultural and academic backgrounds. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 457–458.

意见 2：

此外，还有一些问题没有修改好：

我提到的在讨论中使用小标题的建议，是希望能基于讨论内容增加实质性的小标题，比如理论贡献可以分几个小标题深入讨论。修改后的讨论依然缺乏理论深度，反而很大篇幅讨论研

究局限。

回应：

非常感谢您的意见。在上一轮修改过程中我们并没能很好地理解您所提出增加讨论小标题的审稿意见的真正含义，对此我们感到十分抱歉。在这一轮修改过程中，我们对原有讨论进行了重新的梳理、组织和扩充，将讨论内容框定为三大部分，即理论贡献、实践启示、以及研究局限与未来展望。但为了能够通过小标题一目了然的说明每一部分讨论的核心要旨，我们对前两部分的讨论内容进行了概括，并以其大意将其命名为“被拒斥的算法管理”与“通向智能管理的可行之路”。

在第一小节中，我们强调了本研究的重要理论意义立足于“算法作为领导者”的范式转变。其次，我们重点讨论了本研究对所关切研究问题的主要结论，即我们发现了在职场决策中人们对算法管理的厌恶倾向，并且将厌恶倾向从认知-情感-行为意图三个不同层面测量。最后，我们提出，本研究主要结论的终极指向不是让人有理由拒斥算法，而是警醒人们思考算法管理背景下的人机关系。而在第二小节中，我们整理了本研究的其他重要发现，将实验 2 和 3 验证的透明性的中介机制、实验 4 发现的拟人化的调节作用、以及实验 4 中发现对算法熟悉、了解程度的正向影响总结提升为通向智能管理的三条实践启示。透明性的中介作用促进可解释性人工智能的未来发展，拟人化的调节作用提供算法研发的参考思路，对算法了解和熟悉程度的意外发现指向了提升智能职场中员工算法素养的必然需求。在第三小节，我们沿用了“研究局限与未来展望”这一二级标题，讨论了本研究尚存在的种种不足之处，并逐一指出了未来可供参考的研究方向。另外，我们在最前端总结研究结果的部分也指出了实验 4 中发现被试对算法的了解、熟悉程度正向影响积极态度的相关研究结果。希望通过对原本摘要的深化和重构，修改后的总讨论能够达到发表要求。

具体内容如下：

6 总讨论

本研究探索了在职场应用场景下，人们对人类决策者和算法决策者的态度是否存在差异，并在此基础上探讨了造成如此差异的内在机制和边界条件。通过 4 项递进实验，我们发现相对于人类决策者，当算法走上人力资源管理岗位时会引发人们更加苛刻的评价，这是由于算法决策（相比于人类）通常被人们感知为是更不透明的，并且这一差异受到算法拟人化的调节。具体而言，通过为不同被试呈现由人类或算法完成的同样的人力决策并测量其态度，我们发现人们更不允许、更不喜欢、更不愿意利用算法所做的决策，在认知、情感、行为三个不同维度上均会表现出对算法决策排斥，且这一差异具有一定程度的稳健性（实验 1–4）。通过测量被试感知到的决策透明性（实验 2）以及操纵算法决策的可解释性（实验 3），我们进一步发现感知透明性是造成人们对不同决策主体（人类 vs.算法）产生不同决策态度的心理机制，即相比于人类决策，算法进行同样的决策时，人们认为其决策过程是更不透明、更不可理解的，因此造成了对算法的回避。通过对算法拟人化程度的操纵（实验 4），我们发现了决策主体对决策态度的影响受到拟人化的调节作用。当算法拟人化程度高时，人们对算法决策的态度便会更加宽容。这说明拟人化算法是减少职场中算法厌恶的有效途径之一。并且在分析控制变量的过程中，我们发现了被试对算法的熟悉、了解程度的个体差异能够正向预测其对算法管理的积极态度，这似乎说明提升民众的算法素养也能有效推进智能管理。在研究中我们考察了算法决策在职场中的不同应用场景，包括招聘录用（实验 1）、年终奖分配（实验 2）、简历筛选（实验 3）、绩效考核（实验 4）；并且研究的样本涵盖了不同被试，包括来自 Credamo（实验 2、3、4）、PowerCX（实验 4）的全国范围内被试以及来自某高校的大学生被试（实验 1）。正因此，多样化的实验情境材料和被试选取保障了研究结果的稳健性。

6.1 被拒斥的算法管理

随着工业 4.0 时代的逐步推进，由信息技术赋能的算法越发深入地介入到人们社会生活的方方面面之中，人与自动化机器（泛指计算机算法、人工智能等等一系列解放人类劳动的机器）的关系发生了革命性的范式转变。从原本类似主奴关系的“用户-工具（user-tool）”范式，进展到更为平等的“合作伙伴（partner）”范式，到如今方兴未艾的“下属-领导（subordinate-leader）”关系模式彻底颠覆了原先人们对人与机器关系的认识和理解（Wesche & Sonderegger, 2019）。经典的计算机作为社会行动者（computers as social actors, CASA）范式认为，人们会将计算机和其他先进信息技术视作独立的社会实体，与之的互动也会遵从人类社会的社交规则，而不简单认为其是人类编程的死板呈现,（Nass et al., 1994）。由此引申出的“计算机作为领导者（computers as leaders）”范式强调，自动化算法将成为管理层级结构中的中层领导，负责实现高层管理者与底层员工之间的上传下达（Wesche & Sonderegger, 2019）。在此背景下，对于人们如何看待算法管理的研究是大有裨益的。本研究发现了人们对职场中相关决策的算法使用持较为稳定的排斥和厌恶态度，这与先前研究结果基本一致（Acikgoz et al., 2020; Diab et al., 2011; Nørskov et al., 2020）。一项针对德国民众的代表性调查显示，79%的受访者表示他们对于由算法做决策的想法感到不安，并且更加倾向于由人类完成决策（Fischer & Peterson, 2018）。并且，研究发现大部分管理者虽然不排斥算法进入到职场决策（聘用、解雇、分配奖金等等）之中，但其普遍认为应该在人机协作中赋予人类更高的决策权重（Haesevoets et al., 2021）。本研究发现，面对人类与算法做出相同的人力决策时，人们依然会更抵触算法管理。这一结论为职场中的算法厌恶提供了新的证据。同时，这一发现提醒研究者和企业管理者们需要重审作为领导的算法与人类之间的关系，并思考如何使算法管理在普通员工的心理层面平滑过渡，如何让人力算法真正被接纳为企业中的一个独立主体。

值得一提的是，本研究通过态度的不同维度以及职场中的不同视角构建出“三维度-两视角”的因变量模型以测定人们对职场中算法使用的厌恶倾向，扩充了先前研究在表征算法厌恶上的片面性。过往研究大多都只测量被试对算法决策或机器人服务的一个反应指标，例如信任（Hoff & Bashir, 2015）、购买意愿（Wien & Peluso, 2021），并试图以此说明人们的算法厌恶倾向。但这很可能是单薄的，人们很有可能出现态度中知行的不一致，比如认可算法优越性的同时排斥使用算法推荐系统（Yeomans et al., 2019）。结合对人类心理活动的认识，我们认为算法厌恶这种心理现象至少可以从认知、情感、行为（意图）三个维度加以理解和考察（参考态度 ABC 理论，Breckler, 1984）。因此，本研究选取了可容许性、喜爱程度、利用意愿三个变量作为上述三个维度的表征。可容许性从认知维度反映了人们对决策主体进行决策活动的合法性认识，即一般地认为某一决策主体具有行使决策的能力、权利、资格；喜爱是人接应外物而产生的一种积极情感，也是人能在短时间内对外物做出的直觉判断（Bartneck et al., 2009），其反映了人们在情感上对所面对之事物的倾向性；利用意愿则考察了被试基于假想的管理者视角的行为意图。另外，为力求更加全面立体地还原人们对算法管理的态度，除了上述三维度之外，本研究所选取的因变量还兼顾职场中“员工-领导（employee-employer）”的双重视角（Cummins, 1998; Gigerenzer & Hug, 1992）。其中可容许性和喜爱程度是从员工的角度来看待算法，而利用意愿则是要求被试想象自己作为企业负责人对算法的接受性反应。这一视角上的差异或许说明了为什么被试在利用意愿上的算法厌恶在实验 1 和实验 4 中并没有得到复制。或许原因在于被试样本的年轻化（其中很大一部分为学生被试），导致其并没有作为企业负责人的生活经验和体会，甚至可能没有与企业负责人交际的经历，很难想象这一身份的选择倾向性。

诚然，本研究结论证明人们存在一种对职场中算法使用的普遍厌恶倾向，但这并不意味着企业因此就要放弃数字化进程。原本为提升员工福祉和企业绩效而发明设计的算法技术（Benlian et al., 2022），虽不幸因其不透明的属性而从解放人反过来成为奴役人的超级工具，

最终招致怀疑和拒斥（Jussupow et al., 2020）。但这都不足以让人类选择一条因噎废食的错误道路，即忽视算法管理的潜在优势以及对弊端的改进就仓促地决定放弃研发和使用。并且，在对算法管理的优化升级过程中，研究者和一线管理人员不仅需要琢磨如何设计出更为人接纳、助人成就的算法决策系统（Simth & Shum, 2018），仍然需要关注在算法管理背景之下人机关系、人人关系会出现何种程度的变化，并对这些变化进行进一步的评估和取舍。总而言之，本文实际上是在敲响警钟，面对这一推动管理革新的重要角色，我们必须正确、准确地审视其“利弊兼具”的双刃剑特性，扬长而避短，发挥出其最大的效能。

6.2 通向智能管理的可行之路

从实践价值上来说，本研究致力于探索接纳算法管理的关键因素，从而推动人力决策的自动化、智能化。结合四个子实验的发现，本研究提出三条通向智能管理的可行道路。

第一，提高算法决策的透明性，开发可解释性人工智能。实验 2 发现，人们对算法和人类两种不同决策主体的态度差异根源在于人们认为算法的决策过程相较于人类更加不透明、更加不可理解。实验 3 发现，当打开“黑箱”，提升算法决策的可解释性，能够有效改善人们对其原本抱持的消极态度。总而言之，有关于透明性的发现既作为中介机制解释了人们对算法管理厌恶态度的一种可能来源，同时暗示出一个提升智能管理效能的可行方案。一般而言，算法的透明性包含两方面意指，其一是指作为决策主体的算法呈现其自身决策的过程和依据，使原本不可知的过程可知，另一方面则是向观测者展现算法运算的底层规则和逻辑（Confalonieri et al., 2021; Leichtmann et al., 2023）。因此，所谓提升算法透明性，在技术层面须通过开发可解释性人工智能（Explainable Artificial Intelligence, XAI），使得使用者能够理解、清楚 AI 算法做决策之原委，使原本密不透光之“黑箱（black box）”转变为澄明剔透之“玻璃箱（glass box）”（Rai, 2019）。

第二，设计拟人化的管理类算法。实验 4 发现，对算法进行姓名、表达风格的浅层拟人化处理能够有效改善人们对其的厌恶态度，即相比于拥有机械名字、第三人称表述的算法管理者，人们更容许、更喜欢由拥有类人名字、第一人称表述的算法管理者做出绩效考核决策。这与先前关于拟人化带来优势效应和积极体验的研究基本保持一致（Han, 2021; Natarajan & Gombolay, 2020; Yuan & Dennis, 2019）。基于研究结论，本研究倡导算法管理系统的设计方可采用拟人化的形式，将“冷酷无情”的算法升温，以提高人们对其进入决策领域的接纳程度。

第三，提高民众的算法素养。在实验 4 中，本研究发现对算法的熟悉和了解程度两个控制变量对因变量有正向影响，即对算法越熟悉越了解的被试，其对算法管理的态度也更为积极。这一发现可以用单纯曝光效应（mere exposure effect）进行解释，即只要将某些外部信息反复呈现给人，人们对其的喜爱程度就有可能会提高（Zajonc, 1968）。此发现说明，如果想要提高算法管理在群众之中的接受程度，或许可以通过提升人们对算法的熟悉和了解程度。更广义而言，则需要提高民众的算法素养（algorithmic literacy），即用户围绕算法产生的意识、知识、想象、策略和技能（Swart, 2021）。生活于自动化、信息化、智能化的社会中，人们需要提高自身的知识见闻以应对须臾不可离的算法。同时告知管理者们应当致力于培养员工的算法意识、算法身份（参考 IT identity，Craig et al., 2019），从而能够更好地推动现代管理的智能化。

6.3 研究局限与未来展望

当然，本研究仍存在一定的局限性。第一，在因变量选取上虽然涵盖了人们对算法反应的认知、情感、行为三维度和管理者-下属两个视角，但也只是关注了每个维度其中的一个方面，即对决策的可容许性、喜爱程度和利用意愿，仍有许多其他的重要因素有待后续实验测量：例如对于组织承诺极为重要的信任（Logg et al., 2019）、涉及决策程序与结果的公平感知（Schoeffer et al., 2022）、由恐怖谷效应引申而来的怪诞感（Mende et al., 2019）、当决

策失误或不当后可能出现的指责（Malle et al., 2016）、惩罚行为（Lokhorst & van den Hoven, 2011）等。另外，利用意愿仅仅能代表人们自我报告出的对再次使用该种决策者的行为意图，缺少相对客观和准确的测量，不具有较好的生态效度。并且，人们很可能受困于口是心非、知行不一，在真实的管理场景中做出与实验室不相一致的行为反应。因此，未来研究可以将本研究揭示的效应放入真实的组织环境中再做验证，采用现场观察等方式记录被试对不同类型管理者（涉及人与不同形式的算法）的真实反映和反馈。

其次，本研究仅仅证明了赋予算法以人类名称和拟人化的言辞表达（例如，第一人称）这种最表层的拟人化方法的有效性。一方面这是因为言辞表达和名称使用是实际应用中最为简单的拟人化方法（例如苹果公司的智能语音助手 Siri、小米公司的语音交互引擎小爱同学早已投入生产和使用），最能够得到普遍的应用；另一方面也是因为外观、动作等更加深度的拟人化对人们态度的影响更加复杂，甚至可能出现恐怖谷效应（uncanny valley effect，Laakasuo et al., 2021; Mori, 1970; Mori et al., 2012）、身份威胁（Yoggeeswaran et al., 2016; Zlotowski et al., 2017）等适得其反的干扰。因此，未来研究可以更加深入探讨不同类型、不同程度的拟人化能对人们对于职场算法的态度产生如何的影响。

当然，职场中算法厌恶可能仍存在其他解释机制和边界条件。在本研究中，我们着重探究了感知透明性的中介作用。可实际上人们的认知是复杂多维的，其他因素也可能成为中介。例如，更少的自由意志（许丽颖 等, 2022）、更低的心智感知水平（Bigman & Gray, 2018）、造成的人类独特性威胁（human uniqueness threat）（Ferrari et al., 2016）都可能是职场决策中算法厌恶的潜在原因。未来的研究可以更加细致地去考察这些可能的变量，并对这些影响因素加以综合比较，以便更彻底地认识人对算法与对人类的反应差异的复杂机制。同样，影响人们的算法接纳度的边界条件绝不止拟人化。对于机器接纳度的探索和降低算法厌恶的努力至少可以从三方面来理解，即人类自身特质、机器自有属性和人机互动模式（许丽颖，喻丰,2020）。因此，算法使用者的个体差异、算法及其实体给使用者造成的不同心理感知（如温暖/能力，Fiske et al., 2002）、人−算法协作的权重模式都可能会潜在地影响对算法管理接纳度，这仍有待后续研究不断深入探索。

参考文献：

Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment, 28*(4), 399–416.

Bartneck, C., Kulic, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics, 1*(1), 71–81.

Benlian, A., Wiener, M., Cram, W., Krasnova, H., Maedche, A., Möhlmann, M., Recker, J., & Remus, U. (2022). Algorithmic management: Bright and dark sides, practical implications, and research opportunities. *Business & Information Systems Engineering, 64*(6), 825–839.

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*, 21–34.

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology, 47*(6), 1191–1205.

Confalonieri, R., Coba, L., Wagner, B., & Besold, T. R. (2021). A historical perspective of explainable Artificial Intelligence. *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery, 11*(1), e1391.

Craig, K., Thatcher, J. B., Grover, V. (2019). The IT identity threat: A conceptual definition and operational measure. Journal of Management Information Systems, 36(1), 259–288.

Cummins, D. (1998). Social norms and other minds: The evolutionary roots of higher cognition. In D. D. Cummins & C. Allen (Eds), *The evolution of mind* (pp. 30–50). New York: Oxford University Press.

Diab, D. L., Pui, S. Y., Yankelevich, M., & Highhouse, S. (2011). Lay perceptions of selection decision aids in US and non‐US samples. *International Journal of Selection and Assessment, 19*(2), 209–216.

Ferrari, F., Paladino, M., & Jetten, J. (2016). Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics*, *8*(2), 287–302.

Fischer, S., & Peterson, T. (2018). *Was Deutschland über Algorithmen weiß und denkt: Ergebnisse einer repräsentativen Bevölkerungsumfrage [What Germany knows and think about algorithms: Results of a representative survey]*. Gütersloh, Germany: Bertelsmann Stiftung.

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. *Journal of Personality and Social Psychology, 82*(6), 878–902.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating and perspective change. *Cognition, 43*(2), 127–171.

Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior, 119*, 106730.

Han, M. (2021). The impact of anthropomorphism on consumers' purchase decision in chatbot commerce. *Journal of Internet Commerce, 20*(1), 46–65.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 57*(3),407–434.

Jussuow, E., Benbasat, I., Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. *In Proceedings of the 28th European Conference on Information Systems (ECIS)*, 1–16.

Laakasuo, M., Palomäki, J., & Köbis, N. (2021). Moral uncanny valley: A robot's appearance moderates how its decisions are judged. *International Journal of Social Robotics, 13*(7), 1679–1688.

Leichtmann, B., Humer, C., Hinterreiter, A., Streit, M., & Mara, M. (2023). Effects of explainable artificial intelligence on trust and human behavior in a high-risk decision task. *Computers in Human Behavior, 139*, 107539.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes,151*, 90–103.

Lokhorst, G. J., & van den Hoven, J. (2011). *Responsibility for military robots*. In P. Lin, K. Abeney, & George A. Bekey (Eds.). *Robot ethics: the ethical and social implications of robotics* (pp. 145–156). Cambridge, MA: The MIT Press.

Malle, B. F., Scheutz, M., Forlizzi, J., & Voiklis, J. (2016). Which robot am I thinking about? The impact of action and appearance on people's evaluations of a moral robot. *2016 11th ACM/IEEE international conference on human–robot interaction (HRI)*, 125–132.

Mende, M., Scott, M., Van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising. *Journal of Marketing Research*, *56*(4), 535–556.

Mori, M. (1970). The uncanny valley. *Energy, 7*, 33–35.

Mori, M., MacDorman, K. F., Kageki, N. (2012). The uncanny valley [From the field]. *IEEE Robotics & Automation Magazine, 19*(2), 98–100.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72–78.

Natarajan, M., & Gombolay, M. (2020). Effects of anthropomorphism and accountability on trust in human robot interaction. *2020 15th ACM/IEEE International Conference on Human‐Robot Interaction (HRI)*, 33–42.

Nørskov, S., Damholdt, M., Ulhøi, J., Jensen, M., Ess, C., & Seibt, J. (2020). Applicant fairness perceptions of a

robot‐mediated job interview: A video vignette‐based experimental survey. *Frontiers in Robotics and AI, 7*, 586263.

Rai, A. (2020). Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science, 48*(1), 137–141.

Schoeffer, J., Machowski, Y., & Kuehl, N. (2022). Perceptions of fairness and trustworthiness based on explanations in human vs. automated decision-making. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 1095-1102.

Smith, B., & Shum, H. (2018). *The future computed: Artificial intelligence and its role in society*. Independently Published By Microsoft.

Swart, J. (2021). Experiencing algorithms: How young people understand, feel about, and engage with algorithmic news selection on social media. *Social Media Society, 7*(2), 205630512110088.

Wesche, J. S., & Sonderegger, A. (2019). When computers take the lead: The automation of leadership. *Computers in Human Behavior, 101*, 197–209

Wien, A., & Peluso, A. (2021). Influence of human versus AI recommenders: The roles of product type and cognitive processes. *Journal of Business Research, 137*, 13–27.

Xu, L., & Yu, F. (2020). Factors that influence robot acceptance. *Chinese Science Bulletin, 65*(6), 496-510.

[许丽颖, 喻丰. (2020). 机器人接受度的影响因素. *科学通报, 65*(6), 496–510.]

Xu, L., Yu, F., & Peng, K. (2022). Algorithmic discrimination causes less desire for moral punishment that human discrimination. *Acta Psychological Sinica, 54*(9), 1076-1092.

[许丽颖, 喻丰, 彭凯平. (2022). 算法歧视比人类歧视引起更少道德惩罚欲. *心理学报, 54*(9), 1076–1092.]

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making, 32*(4), 403–414.

Yogeeswaran, K., Złotowski, 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.

Yuan, L., & Dennis, A. (2019). Acting like humans? Anthropomorphism and consumer's willingness to pay in electronic commerce. *Journal of Management Information Systems, 36*(2), 450–477.

Zajonc, R. B. (1968). Attitude effects of mere exposure. *Journal of Personality and Social Psychology, 9*(2), 1–27.

Złotowski, 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.

意见 3：作者根据《心理学报英文摘要写作注意事项》重新润色后的英文摘要依然不专业。举个例子，对于这句话：Experiment 2 was similar to Experiment 1, the only difference was an additional measurement of the mediating role of perceived transparency. 一般情况下，不会把 2 个独立的句子用逗号间隔，组成 1 句话。

回应：

非常感谢您的建议。首先，您在审稿意见中指出的错误的确是我们写作时的疏忽，已在当前摘要的第二段进行了修改。其次，为了完善英文摘要的写作，我们参考了一些发表于《心理学报》的优秀论文的英文摘要，并根据《心理学报英文摘要写作注意事项》的要求重新进行了英文摘要的写作。再邀请四位英文较好的专业人士对内容以及语言进行细致入微的把关和修改。四位分别是：①来自南卡罗来纳大学心理学博士后（澳门大学心理学博士）、②来自香港浸会大学心理学博士后（香港理工大学应用社会科学博士）、③国内高校博士后（清华大学心理学博士）以及④具有英国留学经历（利兹大学）的博士研究生。目前整体四段结构完全参照学报规范，字数为 551 词符合学报要求，各自然段内容和字数均符合学报要求。

第一段主要论述该研究的背景并提出研究问题，第二段介绍各个子研究的研究方法，第三段展现四个递进研究的主要结果，最后一段对文章的贡献、价值、意义进行了讨论和延伸。希望通过上述工作，本研究英文摘要的质量和专业性能得以提升达到发表要求。此版英文摘要呈现如下，敬请审稿专家再次审阅：

Perceived opacity leads to algorithm aversion in the workplace

Abstract

With algorithms standing out and influencing every aspect of human society, people's attitudes toward algorithmic invasion have become a vital topic to be discussed. Recently, algorithms as alternatives and enhancements to human decision-making have become ubiquitously applied in the workplace. Despite algorithms offering numerous advantages, such as vast data storage and anti-interference performance, previous research has found that people tend to reject algorithmic agents across different applications. Especially in the realm of human resources, the increasing utilization of algorithms forces us to focus on users' attitudes. Thus, the present study aimed to explore public attitudes toward algorithmic decision-making and probe the underlying mechanism and potential boundary conditions behind the possible difference.

To verify our research hypotheses, four experiments (N = 1211) were conducted, which involved various kinds of human resource decisions in the daily workplace, including resume screening, recruitment and hiring, allocation of bonuses, and performance assessment. Experiment 1 used a single-factor, two-level, between-subjects design. 303 participants were randomly assigned to two conditions (agent of decision-making: human versus algorithm) and measured their permissibility, liking, and willingness to utilize the agent. Experiment 1 was designed to be consistent with experiment 2. The only difference was an additional measurement of perceived transparency to test the mediating role. Experiment 3 aimed to establish a causal chain between the mediator and dependent variables by manipulating the perceived transparency of the algorithm. In experiment 4, a single-factor three-level between-subjects design (non-anthropomorphism algorithm versus anthropomorphism algorithm versus human) was utilized to explore the boundary condition of this effect.

As anticipated, the present research revealed a pervasive algorithmic aversion across diverse organizational settings. Specifically, we conceptualized algorithm aversion as a tripartite framework encompassing cognitive, affective, and behavioral dimensions. We found that compared with human managers, participants demonstrated significantly lower permissibility (Experiments: 1, 2, and 4), liking (Experiments: 1, 2, and 4), and willingness to utilize (Experiment 2) algorithmic management. And the robustness of this result was demonstrated by the diversity of our scenarios and samples. Additionally, this research discovered perceived transparency as an interpretation mechanism explaining participants' psychological reactions to different decision-making agents. That is to say, participants were opposed to algorithmic management because they thought its decision processes were more incomprehensible and inaccessible than humans (noted in Experiment 2). Addressing this "black box" phenomenon, experiment 3 showed that providing more information and principles about algorithmic management positively influenced participants' attitudes. Crucially, the result also demonstrated the moderating effect of anthropomorphism. The result showed that participants exhibited greater permissibility and liking for the algorithm with human-like characteristics, such as a human-like name and communication style, over more than a mechanized form of the algorithm. This observation underlined the potential of anthropomorphism to ameliorate resistance to algorithmic

management.

These results bridge the gap between algorithmic aversion and decision transparency from the social-psychological perspective. Firstly, the present research establishes a three-dimensional (cognitive, affective, and behavioral) dual-perspective (employee and employer) model to elucidate the negative responses toward algorithmic management. Secondly, it reveals that perceived opacity acts as an obstacle to embracing algorithmic decision-making. This finding lays the theoretical foundation of Explainable Artificial Intelligence (XAI) which is conceptualized as a "glass box". Ultimately, the study highlights the moderating effect of anthropomorphism on algorithmic aversion. This suggests that anthropomorphizing algorithms could be a feasible approach to facilitate the integration of intelligent management systems.

意见 4：不知道修改后的中介图作者参考了哪篇期刊论文的画法，一般不会把 p 值和星号放在一起，如"p < 0.001***"，属于信息冗余。

回应：

非常感谢您的建议。很抱歉在上一轮审稿中您只表达不对，而未说明如何不对，但我们未能很好地理解您的修改意见，导致此问题保留到了这一轮审稿之中。针对于您指出的中介图的画法问题，我们发现也有诸多发表于《心理学报》的新近文章采纳我们原始的画法（杨焕, 卫旭华, 2022; 张萱, 刘萍萍, 2023）。但既然您提出了信息赘余问题，我们参考多篇发表于社会心理学领域顶刊 Journal of Personality and Social Psychology 的部分文章（Prinzing et al., 2023; O'Keefe et al., 2023; Yu & Zhang, 2023; Bayraktaroglu et al., 2023; Van Zant et al., 2023; Proudfoot & Kay, 2023; Thürmer & Kunze, 2023; Gray et al., 2022; Carey et al., 2022; Fischer et al., 2022）中关于中介示意图的画法（其共性在于只报告中介效应分析的回归系数β 和显著性水平 *p*），将文中实验 2 部分所涉及到的中介效应分析图修改如下，敬请您再度审阅：

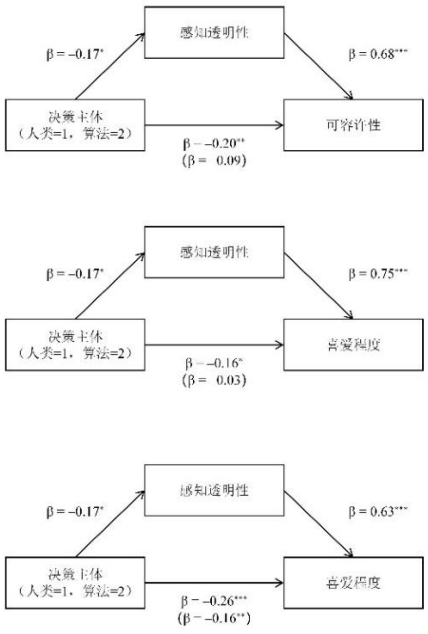


图 2：感知透明性的中介作用

注：*p < 0.05, *p < 0.01, ***p < 0.001

参考文献：

Bayraktaroglu, D., Gunaydin, G., Selcuk, E., Besken, M., & Karakitapoglu-Aygun, Z. (2023). The role of positive

relationship events in romantic attachment avoidance. *Journal of Personality and Social Psychology, 124*(5), 958–970.

Carey, R. M., Stephens, N. M., Townsend, S. S. M., & Hamedani, M. G. (2022). Is diversity enough? Cross-class interactions in college occur less often than expected, but benefit members of lower status groups when they occur. *Journal of Personality and Social Psychology, 123*(5), 889-908.

Fischer, M., Twardawski, M., Strelan, P., & Gollwitzer, M. (2022). Victims need more than power: Empowerment and moral change independently predict victims' satisfaction and willingness to reconcile. *Journal of Personality and Social Psychology, 123*(3), 518-536.

Gray, K., MacCormack, J. K., Henry, T., Banks, E., Schein, C., Armstrong-Carter, E., Abrams, S., & Muscatell, K. A. (2022). The affective harm account (AHA) of moral judgment: Reconciling cognition and affect, dyadic morality and disgust, harm and purity. *Journal of Personality and Social Psychology, 123*(6), 1199-1222.

O'Keefe, P. A., Horberg, E. J., Lee, F., & Dweck, C. S. (2023). Implicit theories of opportunity: When opportunity fails to knock, keep waiting, or start cultivating? *Journal of Personality and Social Psychology, 124*(6), 1146–1173.

Prinzing, M., Le Nguyen, K., & Fredrickson, B. L. (2023). Does shared positivity make life more meaningful? Perceived positivity resonance is uniquely associated with perceived meaning in life. *Journal of Personality and Social Psychology, 125*(2), 345–366.

Proudfoot, D., & Kay, A. C. (2023). Communal expectations conflict with autonomy motives: The western drive for autonomy shapes women's negative responses to positive gender stereotypes. *Journal of Personality and Social Psychology, 124*(1), 1-21.

Thürmer, J. L., & Kunze, F. (2023). Reaction to poor performers in task groups: A model of pro-group intent. *Journal of Personality and Social Psychology, 124*(1), 123-144.

Van Zant, A. B., Kennedy, J. A., & Kray, L. J. (2023). Does hoodwinking others pay? The psychological and relational consequences of undetected negotiator deception. *Journal of Personality and Social Psychology, 124*(5), 1001–1024.

Yu, Y., & Zhang, Y. (2023). The impact of social identity conflict on planning horizons. *Journal of Personality and Social Psychology, 124*(5), 917–934.

杨焕, 卫旭华. (2022). 关系型人力资源管理实践对受益人利他行为的影响：基于道德补偿的视角. *心理学报, 54*(10), 1248–1261.

张萱, 刘萍萍. (2023). 熟悉度促进人们与垃圾分类中的志愿者合作及其作用机制. *心理学报, 55*(8), 1358–1371

审稿人 3 意见：

此轮修改已经较好地回答了之前提出的问题。只有讨论部分理论贡献讨论略显不足，跟前人理论相关很少，建议再深入挖掘一下。鉴于本论文可能存在的实践价值，推荐修改后接受发表。

回应：

非常感谢您的意见。在上一轮修改过程中我们并没能很好地理解您所提出增加讨论小标题的审稿意见的真正含义，对此我们感到十分抱歉。在这一轮修改过程中，我们对原有讨论进行了重新的梳理、组织和扩充，将讨论内容框定为三大部分，即理论贡献、实践启示、以及研究局限与未来展望。但为了能够通过小标题一目了然的说明每一部分讨论的核心要旨，我们对前两部分的讨论内容进行了概括，并以其大意将其命名为“被拒斥的算法管理”与“通向智能管理的可行之路”。

在第一小节中，我们强调了本研究的重要理论意义立足于“算法作为领导者”的范式转变。其次，我们重点讨论了本研究对所关切研究问题的主要结论，即我们发现了在职场决策中人们对算法管理的厌恶倾向，并且将厌恶倾向从认知-情感-行为意图三个方面测量。最后，我们提出，本研究主要结论的终极指向不是让人有理由拒斥算法，而是警醒人们思考算法管理背景下的人机关系。而在第二小节中，我们整理了本研究的其他重要发现，将实验 2 和 3 验证的透明性的中介机制、实验 4 发现的拟人化的调节作用、以及实验 4 中发现对算法熟悉、了解程度的正向影响总结提升为通向智能管理的三条实践启示。透明性的中介作用促进可解释性人工智能的未来发展，拟人化的调节作用提供算法研发的参考思路，对算法了解和熟悉程度的意外发现指向了提升智能职场中员工算法素养的必然需求。在第三小节，我们沿用了“研究局限与未来展望”这一二级标题，讨论了本研究尚存在的种种不足之处，并逐一指出了未来可供参考的研究方向。另外，我们在最前端总结研究结果的部分也指出了实验 4 中发现被试对算法的了解、熟悉程度正向影响积极态度的相关研究结果。希望通过对原本摘要的深化和重构，修改后的总讨论能够达到发表要求。

具体内容如下：

6 总讨论

本研究探索了在职场应用场景下，人们对人类决策者和算法决策者的态度是否存在差异，并在此基础上探讨了造成如此差异的内在机制和边界条件。通过 4 项递进实验，我们发现相对于人类决策者，当算法走上人力资源管理岗位时会引发人们更加苛刻的评价，这是由于算法决策（相比于人类）通常被人们感知为是更不透明的，并且这一差异受到算法拟人化的调节。具体而言，通过为不同被试呈现由人类或算法完成的同样的人力决策并测量其态度，我们发现人们更不允许、更不喜欢、更不愿意利用算法所做的决策，在认知、情感、行为三个不同维度上均会表现出对算法决策排斥，且这一差异具有一定程度的稳健性（实验 1–4）。通过测量被试感知到的决策透明性（实验 2）以及操纵算法决策的可解释性（实验 3），我们进一步发现感知透明性是造成人们对不同决策主体（人类 vs.算法）产生不同决策态度的心理机制，即相比于人类决策，算法进行同样的决策时，人们认为其决策过程是更不透明、更不可理解的，因此造成了对算法的回避。通过对算法拟人化程度的操纵（实验 4），我们发现了决策主体对决策态度的影响受到拟人化的调节作用。当算法拟人化程度高时，人们对算法决策的态度便会更加宽容。这说明拟人化算法是减少职场中算法厌恶的有效途径之一。并且在分析控制变量的过程中，我们发现了被试对算法的熟悉、了解程度的个体差异能够正向预测其对算法管理的积极态度，这似乎说明提升民众的算法素养也能有效推进智能管理。在研究中我们考察了算法决策在职场中的不同应用场景，包括招聘录用（实验 1）、年终奖分配（实验 2）、简历筛选（实验 3）、绩效考核（实验 4）；并且研究的样本涵盖了不同被试，包括来自 Credamo（实验 2、3、4）、PowerCX（实验 4）的全国范围内被试以及来自某高校的大学生被试（实验 1）。正因此，多样化的实验情境材料和被试选取保障了研究结果的稳健性。

6.1 被拒斥的算法管理

随着工业 4.0 时代的逐步推进，由信息技术赋能的算法越发深入地介入到人们社会生活的方方面面之中，人与自动化机器（泛指计算机算法、人工智能等等一系列解放人类劳动的机器）的关系发生了革命性的范式转变。从原本类似主奴关系的“用户-工具（user-tool）”范式，进展到更为平等的“合作伙伴（partner）”范式，到如今方兴未艾的“下属-领导（subordinate-leader）”关系模式彻底颠覆了原先人们对人与机器关系的认识和理解（Wesche & Sonderegger, 2019）。经典的计算机作为社会行动者（computers as social actors, CASA）范式认为，人们会将计算机和其他先进信息技术视作独立的社会实体，与之的互动也会遵从人类社会的社交规则，而不简单认为其是人类编程的死板呈现,（Nass et al., 1994）。由此引

申出的“计算机作为领导者（computers as leaders）”范式强调，自动化算法将成为管理层级结构中的中层领导，负责实现高层管理者与底层员工之间的上传下达（Wesche & Sonderegger, 2019）。在此背景下，对于人们如何看待算法管理的研究是大有裨益的。本研究发现了人们对职场中相关决策的算法使用持较为稳定的排斥和厌恶态度，这与先前研究结果基本一致（Acikgoz et al., 2020; Diab et al., 2011; Nørskov et al., 2020）。一项针对德国民众的代表性调查显示，79%的受访者表示他们对于由算法做决策的想法感到不安，并且更加倾向于由人类完成决策（Fischer & Peterson, 2018）。并且，研究发现大部分管理者虽然不排斥算法进入到职场决策（聘用、解雇、分配奖金等等）之中，但其普遍认为应该在人机协作中赋予人类更高的决策权重（Haesevoets et al., 2021）。本研究发现，面对人类与算法做出相同的人力决策时，人们依然会更抵触算法管理。这一结论为职场中的算法厌恶提供了新的证据。同时，这一发现提醒研究者和企业管理者们需要重审作为领导的算法与人类之间的关系，并思考如何使算法管理在普通员工的心理层面平滑过渡，如何让人力算法真正被接纳为企业中的一个独立主体。

值得一提的是，本研究通过态度的不同维度以及职场中的不同视角构建出“三维度-两视角”的因变量模型以测定人们对职场中算法使用的厌恶倾向，扩充了先前研究在表征算法厌恶上的片面性。过往研究大多都只测量被试对算法决策或机器人服务的一个反应指标，例如信任（Hoff & Bashir, 2015）、购买意愿（Wien & Peluso, 2021），并试图以此说明人们的算法厌恶倾向。但这很可能是单薄的，人们很有可能出现态度中知行的不一致，比如认可算法优越性的同时排斥使用算法推荐系统（Yeomans et al., 2019）。结合对人类心理活动的认识，我们认为算法厌恶这种心理现象至少可以从认知、情感、行为（意图）三个维度加以理解和考察（参考态度 ABC 理论，Breckler, 1984）。因此，本研究选取了可容许性、喜爱程度、利用意愿三个变量作为上述三个维度的表征。可容许性从认知维度反映了人们对决策主体进行决策活动的合法性认识，即一般地认为某一决策主体具有行使决策的能力、权利、资格；喜爱是人接应外物而产生的一种积极情感，也是人能在短时间内对外物做出的直觉判断（Bartneck et al., 2009），其反映了人们在情感上对所面对之事物的倾向性；利用意愿则考察了被试基于假想的管理者视角的行为意图。另外，为力求更加全面立体地还原人们对算法管理的态度，除了上述三维度之外，本研究所选取的因变量还兼顾职场中“员工-领导（employee-employer）”的双重视角（Cummins, 1998; Gigerenzer & Hug, 1992）。其中可容许性和喜爱程度是从员工的角度来看待算法，而利用意愿则是要求被试想象自己作为企业负责人对算法的接受性反应。这一视角上的差异或许说明了为什么被试在利用意愿上的算法厌恶在实验 1 和实验 4 中并没有得到复制。或许原因在于被试样本的年轻化（其中很大一部分为学生被试），导致其并没有作为企业负责人的生活经验和体会，甚至可能没有与企业负责人交际的经历，很难想象这一身份的选择倾向性。

诚然，本研究结论证明人们存在一种对职场中算法使用的普遍厌恶倾向，但这并不意味着企业因此就要放弃数字化进程。原本为提升员工福祉和企业绩效而发明设计的算法技术（Benlian et al., 2022），虽不幸因其不透明的属性而从解放人反过来成为奴役人的超级工具，最终招致怀疑和拒斥（Jussupow et al., 2020）。但这都不足以让人类选择一条因噎废食的错误道路，即忽视算法管理的潜在优势以及对弊端的改进就仓促地决定放弃研发和使用。并且，在对算法管理的优化升级过程中，研究者和一线管理人员不仅需要琢磨如何设计出更为人接纳、助人成就的算法决策系统（Simth & Shum, 2018），仍然需要关注在算法管理背景之下人机关系、人人关系会出现何种程度的变化，并对这些变化进行进一步的评估和取舍。总而言之，本文实际上是在敲响警钟，面对这一推动管理革新的重要角色，我们必须正确、准确地审视其“利弊兼具”的双刃剑特性，扬长而避短，发挥出其最大的效能。

6.2 通向智能管理的可行之路

从实践价值上来说，本研究致力于探索接纳算法管理的关键因素，从而推动人力决策的自动化、智能化。结合四个子实验的发现，本研究提出三条通向智能管理的可行道路。

第一，提高算法决策的透明性，开发可解释性人工智能。实验 2 发现，人们对算法和人类两种不同决策主体的态度差异根源在于人们认为算法的决策过程相较于人类更加不透明、更加不可理解。实验 3 发现，当打开“黑箱”，提升算法决策的可解释性，能够有效改善人们对其原本抱持的消极态度。总而言之，有关于透明性的发现既作为中介机制解释了人们对算法管理厌恶态度的始末，同时暗示出一个提升智能管理效能的可行方案。一般而言，算法的透明性包含两方面意指，其一是指作为决策主体的算法呈现其自身决策的过程和依据，使原本不可知的过程可知，另一方面则是向观测者展现算法运算的底层规则和逻辑（Confalonieri et al., 2021; Leichtmann et al., 2023）。因此，所谓提升算法透明性，在技术层面须通过开发可解释性人工智能（Explainable Artificial Intelligence, XAI），使得使用者能够理解、清楚 AI 算法做决策之原委，使原本密不透光之“黑箱（black box）”转变为澄明剔透之“玻璃箱（glass box）”（Rai, 2019）。

第二，设计拟人化的管理类算法。实验 4 发现，对算法进行姓名、表达风格的浅层拟人化处理能够有效改善人们对其的厌恶态度，即相比于拥有机械名字、第三人称表述的算法管理者，人们更容许、更喜欢由拥有类人名字、第一人称表述的算法管理者做出绩效考核决策。这与先前关于拟人化带来优势效应和积极体验的研究基本保持一致（Han, 2021; Natarajan & Gombolay, 2020; Yuan & Dennis, 2019）。基于研究结论，本研究倡导算法管理系统的设计方可采用拟人化的形式，将“冷酷无情”的算法升温，以提高人们对其进入决策领域的接纳程度。

第三，提高民众的算法素养。在实验 4 中，本研究发现对算法的熟悉和了解程度两个控制变量对因变量的有正向影响，即对算法越熟悉越了解的被试，其对算法管理的态度也更为积极。这一发现可以用简单暴露效应（mere exposure effect）进行解释，即只要将某些外部信息反复呈现给人，人们对其的喜爱程度就有可能会提高（Zajonc, 1968）。这发现说明了如果想要提高算法管理在群众之中的接受程度，或许可以通过提升人们对算法的熟悉和了解程度。更广义而言，则需要提高民众的算法素养（algorithmic literacy），即用户围绕算法产生的意识、知识、想象、策略和技能（Swart, 2021）。生活于自动化、信息化、智能化的社会中，人们需要提高自身的知识见闻以应对须臾不可离的算法。同时告知管理者们应当致力于培养员工的算法意识、算法身份（参考 IT identity，Craig et al., 2019），从而能够更好地推动现代管理的智能化。

6.3 研究局限与未来展望

当然，本研究仍存在一定的局限性。第一，在因变量选取上虽然涵盖了人们对算法反应的认知、情感、行为三维度和管理者-下属两个视角，但也只是关注了每个维度其中的一个方面，即对决策的可容许性、喜爱程度和利用意愿，仍有许多其他的重要因素有待后续实验测量：例如对于组织承诺极为重要的信任（Logg et al., 2019）、涉及决策程序与结果的公平感知（Schoeffer et al., 2022）、由恐怖谷效应引申而来的怪诞感（Mende et al., 2019）、当决策失误或不当后可能出现的指责（Malle et al., 2016）、惩罚行为（Lokhorst & van den Hoven, 2011）等。另外，利用意愿仅仅能代表人们自我报告出的对再次使用该种决策者的行为意图，缺少相对客观和准确的测量，不具有较好的生态效度。并且，人们很可能受困于口是心非、知行不一，在真实的管理场景中做出与实验室不相一致的行为反应。因此，未来研究可以将本研究揭示的效应放入真实的组织环境中再做验证，采用现场观察等方式记录被试对不同类型管理者（涉及人与不同形式的算法）的真实反映和反馈。

其次，本研究仅仅证明了赋予算法以人类名称和拟人化的言辞表达（例如，第一人称）这种最表层的拟人化方法的有效性。一方面这是因为言辞表达和名称使用是实际应用中最为

简单的拟人化方法（例如苹果公司的智能语音助手 Siri、小米公司的语音交互引擎小爱同学早已投入生产和使用），最能够得到普遍的应用；另一方面也是因为外观、动作等更加深度的拟人化对人们态度的影响更加复杂，甚至可能出现恐怖谷效应（uncanny valley effect，Laakasuo et al., 2021; Mori, 1970; Mori et al., 2012）、身份威胁（Yoggeeswaran et al., 2016; Zlotowski et al., 2017）等适得其反的干扰。因此，未来研究可以更加深入探讨不同类型、不同程度的拟人化能对人们对于职场算法的态度产生如何的影响。

当然，职场中算法厌恶可能仍存在其他解释机制和边界条件。在本研究中，我们着重探究了感知透明性的中介作用。可实际上人们的认知是复杂多维的，其他因素也可能成为中介。例如，更少的自由意志（许丽颖 等, 2022）、更低的心智感知水平（Bigman & Gray, 2018）、造成的人类独特性威胁（human uniqueness threat）（Ferrari et al., 2016）都可能是职场决策中算法厌恶的潜在原因。未来的研究可以更加细致地去考察这些可能的变量，并对这些影响因素加以综合比较，以便更彻底地认识人对算法与对人类的反应差异的复杂机制。同样，影响人们的算法接纳度的边界条件绝不止拟人化。对于机器接纳度的探索和降低算法厌恶的努力至少可以从三方面来理解，即人类自身特质、机器自有属性和人机互动模式（许丽颖, 喻丰, 2020）。因此，算法使用者的个体差异、算法及其实体给使用者造成的不同心理感知（如温暖/能力，Fiske et al., 2002）、人-算法协作的权重模式都可能会潜在地影响对算法管理接纳度，这仍有待后续研究不断深入探索。

讨论中所用全部参考文献：

Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment, 28*(4), 399–416.

Bartneck, C., Kulic, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics, 1*(1), 71–81.

Benlian, A., Wiener, M., Cram, W., Krasnova, H., Maedche, A., Möhlmann, M., Recker, J., & Remus, U. (2022). Algorithmic management: Bright and dark sides, practical implications, and research opportunities. *Business & Information Systems Engineering, 64*(6), 825–839.

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*, 21–34.

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology, 47*(6), 1191–1205.

Confalonieri, R., Coba, L., Wagner, B., & Besold, T. R. (2021). A historical perspective of explainable Artificial Intelligence. *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery, 11*(1), e1391.

Craig, K., Thatcher, J. B., Grover, V. (2019). The IT identity threat: A conceptual definition and operational measure. Journal of Management Information Systems, 36(1), 259–288.

Cummins, D. (1998). Social norms and other minds: The evolutionary roots of higher cognition. In D. D. Cummins & C. Allen (Eds), *The evolution of mind* (pp. 30–50). New York: Oxford University Press.

Diab, D. L., Pui, S. Y., Yankelevich, M., & Highhouse, S. (2011). Lay perceptions of selection decision aids in US and non‐US samples. *International Journal of Selection and Assessment, 19*(2), 209–216.

Ferrari, F., Paladino, M., & Jetten, J. (2016). Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics, 8*(2), 287–302.

Fischer, S., & Peterson, T. (2018). *Was Deutschland über Algorithmen weiß und denkt: Ergebnisse einer repräsentativen Bevölkerungsumfrage [What Germany knows and think about algorithms: Results of a representative survey]*. Gütersloh, Germany: Bertelsmann Stiftung.

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. *Journal of Personality and Social Psychology, 82*(6), 878–902.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating and perspective change. *Cognition, 43*(2), 127–171.

Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior, 119*, 106730.

Han, M. (2021). The impact of anthropomorphism on consumers' purchase decision in chatbot commerce. *Journal of Internet Commerce, 20*(1), 46–65.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 57*(3),407–434.

Jussuow, E., Benbasat, I., Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. *In Proceedings of the 28th European Conference on Information Systems (ECIS)*, 1–16.

Laakasuo, M., Palomäki, J., & Köbis, N. (2021). Moral uncanny valley: A robot's appearance moderates how its decisions are judged. *International Journal of Social Robotics, 13*(7), 1679–1688.

Leichtmann, B., Humer, C., Hinterreiter, A., Streit, M., & Mara, M. (2023). Effects of explainable artificial intelligence on trust and human behavior in a high-risk decision task. *Computers in Human Behavior, 139*, 107539.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes,151*, 90–103.

Lokhorst, G. J., & van den Hoven, J. (2011). *Responsibility for military robots*. In P. Lin, K. Abeney, & George A. Bekey (Eds.). *Robot ethics: the ethical and social implications of robotics* (pp. 145–156). Cambridge, MA: The MIT Press.

Malle, B. F., Scheutz, M., Forlizzi, J., & Voiklis, J. (2016). Which robot am I thinking about? The impact of action and appearance on people's evaluations of a moral robot. *2016 11th ACM/IEEE international conference on human–robot interaction (HRI)*, 125–132.

Mende, M., Scott, M., Van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising. *Journal of Marketing Research, 56*(4), 535–556.

Mori, M. (1970). The uncanny valley. *Energy, 7*, 33–35.

Mori, M., MacDorman, K. F., Kageki, N. (2012). The uncanny valley [From the field]. *IEEE Robotics & Automation Magazine, 19*(2), 98–100.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72–78.

Natarajan, M., & Gombolay, M. (2020). Effects of anthropomorphism and accountability on trust in human robot interaction. *2020 15th ACM/IEEE International Conference on Human‐Robot Interaction (HRI)*, 33–42.

Nørskov, S., Damholdt, M., Ulhøi, J., Jensen, M., Ess, C., & Seibt, J. (2020). Applicant fairness perceptions of a robot‐mediated job interview: A video vignette‐based experimental survey. *Frontiers in Robotics and AI, 7*, 586263.

Rai, A. (2020). Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science, 48*(1), 137–141.

Schoeffer, J., Machowski, Y., & Kuehl, N. (2022). Perceptions of fairness and trustworthiness based on explanations in human vs. automated decision-making. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 1095-1102.

Smith, B., & Shum, H. (2018). *The future computed: Artificial intelligence and its role in society*. Independently

Published By Microsoft.

Swart, J. (2021). Experiencing algorithms: How young people understand, feel about, and engage with algorithmic news selection on social media. *Social Media Society, 7*(2), 205630512110088.

Wesche, J. S., & Sonderegger, A. (2019). When computers take the lead: The automation of leadership. *Computers in Human Behavior, 101*, 197–209

Wien, A., & Peluso, A. (2021). Influence of human versus AI recommenders: The roles of product type and cognitive processes. *Journal of Business Research, 137*, 13–27.

Xu, L., & Yu, F. (2020). Factors that influence robot acceptance. *Chinese Science Bulletin, 65*(6), 496-510.

[许丽颖, 喻丰. (2020). 机器人接受度的影响因素. *科学通报, 65*(6), 496–510.]

Xu, L., Yu, F., & Peng, K. (2022). Algorithmic discrimination causes less desire for moral punishment that human discrimination. *Acta Psychological Sinica, 54*(9), 1076-1092.

[许丽颖, 喻丰, 彭凯平. (2022). 算法歧视比人类歧视引起更少道德惩罚欲. *心理学报, 54*(9), 1076–1092.]

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making, 32*(4), 403–414.

Yogeeswaran, K., Złotowski, 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.

Yuan, L., & Dennis, A. (2019). Acting like humans? Anthropomorphism and consumer's willingness to pay in electronic commerce. *Journal of Management Information Systems, 36*(2), 450–477.

Zajonc, R. B. (1968). Attitude effects of mere exposure. *Journal of Personality and Social Psychology, 9*(2), 1–27.

Złotowski, 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.

第三轮

编委意见：

我认真看了一下外审的意见和作者的修改稿件，鉴于目前三位评审中，两位外审同意发表，因此同意提交主编进行最终审核。

主编意见：同意发表。