

Ethical behavior recommendations proposed by AI supervisors are less likely to be followed

Authors: Xu Liying, Zhao Yijun, Yu Feng, Yu Feng

Date: 2024-09-04T00:00:00+00:00

Abstract

The rapid development of Artificial Intelligence (AI) technology has triggered tremendous transformations in organizations, with AI assuming supervisory roles that can directly influence employee behavior. Six progressive scenario experiments (N = 1642) sought to investigate differential reactions to ethical behavior recommendations proposed by AI versus human supervisors, along with the underlying psychological mechanisms and boundary conditions. Results revealed that, compared to human supervisors, individuals exhibited lower compliance with ethical behavior recommendations from AI supervisors (Experiments 1a-5). This was attributable to lower evaluation apprehension in interactions with AI supervisors (Experiments 2-3). Moreover, compliance with AI supervisors' ethical recommendations increased when individuals had stronger anthropomorphism tendencies or when the AI supervisor was more anthropomorphized (Experiments 4-5). The findings contribute to a better understanding of reactions to AI supervisors in organizational contexts, illuminate limitations of AI supervisors in ethical guidance domains, and provide practical references and improvement strategies for deploying AI leadership in organizational management.

Full Text

Employees Adhere Less to Moral Behavior Advice from AI Supervisors than from Human Supervisors

XU Liying, ZHAO Yijun, YU Feng

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

Abstract

The rapid development of Artificial Intelligence (AI) technology has triggered profound transformations in organizations, with AI assuming supervisory roles

that directly influence employee behavior. Through six sequential scenario experiments ($N = 1642$), this research investigates differences in how people respond to moral behavior advice proposed by AI versus human supervisors, exploring the underlying psychological mechanisms and boundary conditions. Results demonstrate that individuals exhibit significantly lower compliance with moral behavior advice from AI supervisors compared to human supervisors (Experiments 1a–5). This effect is mediated by lower evaluation apprehension in interactions with AI supervisors (Experiments 2–3). Furthermore, compliance with AI supervisors' moral advice increases when individuals have stronger anthropomorphic tendencies or when the AI supervisor is more anthropomorphized (Experiments 4–5). These findings enhance understanding of employee reactions to AI supervisors in organizational settings, reveal limitations of AI supervisors in moral guidance domains, and provide practical recommendations for deploying AI leadership in organizational management.

Keywords: AI supervisor, advice compliance, moral behavior advice, evaluation apprehension, anthropomorphism

1. Introduction

1.1 Rejection of Moral Behavior Advice from AI Supervisors

AI's ascension to management positions is no longer uncommon. Leveraging advantages such as vast information processing capacity, objectivity, resistance to interference, and enhanced efficiency, AI applications have introduced unprecedented changes and challenges to management. AI can automate numerous managerial functions including monitoring, goal-setting, and performance management, occupying middle-management positions to make autonomous decisions and provide guidance to subordinates. This transformation has fundamentally altered the power structure between humans and machines, with AI evolving from a tool to a supervisor.

While this innovation yields certain benefits—such as improved decision accuracy, enhanced task performance, increased sense of work meaning, and elevated employee loyalty—empirical evidence suggests people resist such changes. Individuals demonstrate aversion to AI exercising managerial authority, perceiving lower organizational commitment and status under AI management, even characterizing it as “tyranny.” Consequently, when AI managers serve in advisory roles, subordinates may maintain this negative attitude, showing reduced willingness to comply.

Previous research reveals widespread algorithm aversion, where people reject and dislike decisions made by algorithms across domains including medical decisions, daily recommendations, and aesthetic creation. In morally relevant contexts, algorithm aversion is particularly pronounced: people are less willing to use AI for decisions with higher moral relevance, preferring human decision-

makers, and they attribute lower moral standing to identical actions performed by AI versus humans. Moreover, individuals reject AI involvement in promoting or designing public welfare and charitable projects. This demonstrates a clear aversion to AI's encroachment into moral domains.

For organizations, morality represents a crucial characteristic and evaluation criterion, with employee moral behavior serving as a significant driver of organizational development. Moral behaviors—such as organizational citizenship behavior and volunteerism—enhance employees' sense of work meaning and vitality at the individual level while fostering harmonious colleague relationships. At the organizational level, they increase productivity, improve customer satisfaction, and construct fair and just organizational cultures. As large language model technology advances, AI's scope has expanded to include prescriptive tasks like goal-setting and planning. While most research has focused on AI's role in facilitating unethical behavior, this study examines psychological reactions when AI becomes a proponent of moral behavior, directly comparing compliance with moral advice from human versus AI supervisors.

Based on the above, we propose Hypothesis H1: Compared to moral behavior advice from human supervisors, people demonstrate lower compliance with moral behavior advice from AI supervisors.

1.2 The Mediating Role of Evaluation Apprehension

What determines our compliance with advice from different supervisors? Previous research suggests people reject discriminatory instructions from AI managers because they perceive AI as lacking mental capacities—a root cause of many human-AI response differences. According to mind perception theory, the mental capacities people attribute to different entities determine their moral status, and people generally believe AI lacks complete mental faculties. However, this perspective only reveals stable, inherent attribute differences between humans and AI, failing to explain response differences from the perspective of interactive experience. In organizational contexts, individuals' perceptions and experiences of objective situations influence attitudes and behaviors more than objective reality itself. Therefore, this research examines the psychological experience of interaction to explain why people reject moral advice from AI supervisors.

In social interactions, impression management needs lead individuals to follow social norms more closely when under others' observation. Consequently, in any social situation, individuals sense potential judgment from others and worry about being evaluated. This experience is conceptualized as “evaluation apprehension”—concern about receiving negative evaluations or failing to receive positive ones from others. In organizational contexts, evaluation apprehension is common, arising from interactions between supervisors and employees and subsequently influencing behavioral performance.

When AI becomes a direct supervisor, does human-AI interaction elicit this psychological perception? And does it differ from that caused by human super-

visors? Previous research indicates that in human-computer interaction, people treat AI as social actors capable of evoking automated social presence, which can trigger evaluation apprehension. However, studies using behavior tracking and virtual/augmented reality technologies for management find they generate lower evaluation apprehension than human managers. Additionally, people perceive AI chatbots as lacking judgment capabilities and causing no negative social consequences, resulting in lower perceived social judgment and reputational damage during actual interactions. We therefore predict that while AI can trigger evaluation apprehension, the level is substantially lower than that elicited by humans.

How does varying evaluation apprehension affect advice compliance? Evaluation apprehension exerts external pressure on individuals to engage in impression-enhancing behaviors. For example, employees perform organizational citizenship behaviors when they believe their actions will receive positive attention from managers. Simultaneously, evaluation apprehension reduces internal motivation, causing withdrawal and loss of confidence, thereby increasing dependence on advice. Overall, evaluation apprehension both motivates behavior that garners positive social evaluation and increases reliance on others' advice, positively influencing compliance with supervisors' moral behavior advice.

Thus, we propose Hypothesis H2: Evaluation apprehension mediates the effect of supervisor type (human vs. AI) on moral behavior advice compliance.

1.3 The Moderating Role of Anthropomorphism

Evaluation apprehension typically emerges in social situations caused by others' presence, potentially affecting physiology, emotion, behavior, and performance. Since social situations usually involve people, discussing whether AI supervisors elicit evaluation apprehension inevitably involves the issue of AI anthropomorphism. Anthropomorphism refers to attributing human characteristics to non-human entities, manifesting both as an object attribute and as a subjective cognitive tendency.

Previous research shows that anthropomorphizing technological entities effectively reduces aversion toward them. From an object attribute perspective, anthropomorphic AI design reduces psychological distance and enhances social presence. From a subject tendency perspective, individuals high in anthropomorphic tendencies anticipate more pleasant interactions with AI, show higher acceptance, and engage more according to social norms. Overall, whether as an individual tendency or object attribute, anthropomorphism reduces response differences between AI and humans, making reactions to AI more similar to those toward humans. Following this logic, high anthropomorphism should reduce the human-AI distinction and thus diminish compliance differences.

Research on advice adoption provides corroborating evidence. In Judge-Advisor Systems, similarity between advisor and decision-maker positively influences advice compliance. When high similarity exists—whether from group identity

matching, consistent behavioral choices, or coincidental similarities—decision-makers are more likely to follow advice. Since AI is typically viewed as an outgroup, similarity with humans is minimal. However, through anthropomorphism, AI is endowed with more human attributes and characteristics, increasing similarity with humans and thereby enhancing compliance with anthropomorphized AI supervisors' advice.

Accordingly, we propose Hypothesis H3: Anthropomorphism moderates the effect of supervisor type (human vs. AI) on moral behavior advice compliance. Specifically, for individuals low in anthropomorphic tendency, compliance is higher for human versus AI supervisors, while this difference becomes non-significant for those high in anthropomorphic tendency. Regarding AI supervisor characteristics, compliance increases as AI anthropomorphism increases.

1.4 Research Overview

This research employs six sequential experiments to investigate employee reactions to moral behavior advice from AI supervisors, exploring psychological mechanisms and boundary conditions. The experiments involve three quantifiable moral behaviors: donation amounts, volunteer hours, and walking steps. Specific scenarios include environmental volunteer services (Experiment 1a), charitable walking activities (Experiment 1b), donations for colleagues in difficulty (Experiments 2 and 4), disaster relief donations (Experiment 3), and child welfare participation (Experiment 5).

Experiments 1a and 1b test the main hypothesis that compliance with AI supervisors' moral advice is lower than with human supervisors. Experiment 2 examines the mediating role of evaluation apprehension. Experiment 3 manipulates evaluation apprehension to further test whether it causes differences in compliance between human and AI supervisors. Experiment 4 explores individual differences in anthropomorphic tendency as a boundary condition. Finally, Experiment 5 directly manipulates AI supervisor anthropomorphism to validate the moderating effect.

2. Pilot Experiment: Testing for Non-Exploitation in Experimental Scenarios

To ensure that the three moral behavior scenarios (walking for charity, donation campaigns, volunteer services) used in subsequent experiments would not be perceived as exploitative moral coercion, we conducted a pilot experiment examining perceptions of exploitation and autonomy.

2.1.1 Participants

Following previous research, we recruited 50 participants ($M_{age} = 32.63 \pm 9.19$, 57.1% female) for this pilot experiment. All participants provided informed

consent after reading experimental instructions and received compensation upon completion. Participants who passed attention checks were included in the final sample.

2.1.2 Procedure

Participants sequentially read three scenario descriptions involving organizational moral behaviors: a walking activity for disability support, a donation campaign for colleagues in difficulty, and a volunteer service for disadvantaged groups. After each scenario, participants rated perceived exploitation and autonomy using 7-point Likert scales. Exploitation perception was measured with three items (e.g., “To what extent do you consider the company’s walking charity activity coercive?”; Cronbach’s $\alpha = 0.76, 0.90, 0.93$ across scenarios). Autonomy perception was measured with three items (e.g., “To what extent did you feel you had autonomy to express your own ideas during the activity?”; Cronbach’s $\alpha = 0.82, 0.91, 0.90$).

2.2 Results

One-sample t-tests on exploitation perception revealed that ratings for all three scenarios were significantly below the scale midpoint: walking scenario ($M = 1.79, SD = 0.78, t(49) = -19.95, p < 0.001, \text{Cohen's } d = -2.82$), donation scenario ($M = 2.19, SD = 0.98, t(49) = -10.95, p < 0.001, \text{Cohen's } d = -1.55$), and volunteer scenario ($M = 2.26, SD = 1.37, t(49) = -8.93, p < 0.001, \text{Cohen's } d = -1.26$). Participants perceived these scenarios as non-exploitative.

One-sample t-tests on autonomy perception showed ratings significantly above the midpoint for all scenarios: walking ($M = 5.29, SD = 0.98, t(49) = 9.36, p < 0.001, \text{Cohen's } d = 1.32$), donation ($M = 5.12, SD = 1.39, t(49) = 5.68, p < 0.001, \text{Cohen's } d = 0.80$), and volunteer ($M = 5.13, SD = 1.40, t(49) = 5.73, p < 0.001, \text{Cohen's } d = 0.81$). Participants felt they maintained autonomous choice.

2.3 Discussion

This pilot experiment confirmed that employees perceive organizational moral behavior activities as non-exploitative and non-coercive, while maintaining a sense of autonomous choice. This eliminates potential confounds of perceived exploitation and autonomy restriction in subsequent experiments.

Experiment 1a: Initial Test of Compliance Differences

Experiment 1a examined whether compliance with moral behavior advice differs between human and AI supervisors. Using an online scenario experiment, we randomly assigned participants to human or AI conditions and measured their planned volunteer hours before and after receiving supervisory advice.

3.1.1 Participants

Using G*Power 3.1, we calculated that 172 participants were needed for 90% power to detect a medium effect ($d = 0.5$) with $\alpha = 0.05$. We recruited 180 valid participants via Credamo (68 male, 112 female; Mage = 30.05, SD = 8.33) after real-time exclusion of failed attention checks.

3.1.2 Design and Procedure

This single-factor between-subjects experiment randomly assigned participants to human or AI supervisor conditions. All participants imagined being a junior employee at a company organizing weekend environmental volunteer activities in surrounding communities (non-profit, no performance requirements). They first indicated how many of 832 possible annual weekend hours they would volunteer.

Participants then received information about company management and supervisory advice. The human condition described traditional professional management; the AI condition described AI algorithm management. To ensure understanding of the hierarchical relationship, we detailed that the supervisor (Manager Wang or Algorithm CSD300) conducted daily performance evaluations, with all behaviors serving as evaluation criteria. An organizational chart [Figure 1: see original paper] visually reinforced the subordinate-supervisor relationship.

Both groups received identical advice emails: “To promote green development and cultivate employee compassion, we recommend you volunteer 416 hours annually for environmental activities.” After reading the email, participants indicated their final volunteer hour commitment (0-832). An attention check asked participants to identify their supervisor type. Finally, participants reported AI familiarity, AI knowledge, gender, and age.

3.2 Results

Advice compliance was calculated as: $(\text{Final Decision} - \text{Initial Decision}) / (\text{Supervisor's Recommendation} - \text{Initial Decision})$. Four participants whose initial plans matched the recommendation were excluded, leaving 89 in the human group and 87 in the AI group.

Independent samples t-test showed significantly higher compliance in the human group ($M = 0.67$, $SD = 0.53$) than the AI group ($M = 0.34$, $SD = 1.46$), $t(174) = 2.00$, $p = 0.047$, Cohen's $d = 0.30$. ANCOVA controlling for gender, age, AI familiarity, and AI knowledge showed a marginally significant difference, $F(1, 170) = 3.39$, $p = 0.067$, $\eta^2 p = 0.020$.

3.3 Discussion

Experiment 1a initially validated that people preferentially comply with human versus AI supervisors' moral advice, while controlling for AI familiarity

and knowledge. However, it involved only conventional volunteer activities and didn't control for perceived advice rationality or negative emotions. Experiment 1b addresses these limitations using an unconventional moral behavior scenario—donating through exercise steps—while further controlling for rationality and emotional effects.

Experiment 1b: Replication in an Unconventional Moral Behavior Scenario

Experiment 1b aimed to replicate Experiment 1a's findings in an unconventional moral behavior context (walking steps donation) while controlling for perceived advice rationality and negative emotional responses.

4.1.1 Participants

Following Experiment 1a's standards, we recruited 172 university students (Mage = 20.39, SD = 1.62; 56.4% female) via Credamo who completed all measures and passed attention checks.

4.1.2 Design and Procedure

This single-factor between-subjects experiment randomly assigned participants to human or AI conditions. Participants imagined being an employee at a company launching a “One Step, One Love” campaign for National Disability Day. The company pledged to donate 1 yuan per 1,000 steps walked during a 7-day period. Participants initially indicated their planned total steps.

After reading about company management (human expert vs. AI algorithm) and viewing organizational charts, participants received advice: “Every step counts. I recommend walking 14,000 steps daily (98,000 total). Of course, you remain free to choose.” This emphasis on freedom helped exclude exploitation concerns. After reading the email, participants indicated their final step commitment and completed an attention check.

We then measured perceived advice rationality (“To what extent is Manager Zhao's/EPbot's advice reasonable?”) and negative emotions using the PANAS negative affect scale (10 items; $\alpha = 0.896$). Finally, participants reported AI familiarity, AI knowledge, gender, and age.

4.2 Results

Independent samples t-test revealed significantly higher compliance for human supervisors ($M = 0.64$, $SD = 0.45$) than AI supervisors ($M = 0.45$, $SD = 0.58$), $t(170) = 2.39$, $p = 0.018$, Cohen's $d = 0.36$.

ANCOVA controlling for perceived rationality, negative emotions, AI familiarity, AI knowledge, gender, and age maintained a significant main effect, $F(1, 155)$

= 3.90, $p = 0.050$, $^2p = 0.025$. The effect persisted even after controlling for individual differences in rationality perception and emotional experience.

4.3 Discussion

Experiment 1b replicated Experiment 1a's findings and extended them to unconventional moral behavior advice while controlling for rationality and negative emotions. However, both sub-studies revealed the phenomenon without exploring underlying mechanisms. Experiment 2 addresses this by examining evaluation apprehension mediation in a charitable donation context.

Experiment 2: Testing the Mediating Role of Evaluation Apprehension

Experiment 2 expanded moral behavior scenario diversity (charitable donation) to enhance robustness while exploring the mediating mechanism of evaluation apprehension and alternative mediators like mind perception.

5.1.1 Participants

Monte Carlo simulations indicated 250 participants were needed for stable correlations. We recruited 250 valid participants ($Mage = 30.74 \pm 7.64$; 64% female) via Credamo after real-time exclusion of failed attention checks.

5.1.2 Design and Procedure

This single-factor between-subjects experiment randomly assigned participants to human or AI conditions. Participants imagined being a sales team leader who received a 10,000 yuan bonus while a team member faced sudden family hardship. They initially indicated their donation amount (0-10,000 yuan).

After management information (human expert vs. AI algorithm) and organizational charts, participants received advice: "Given Zhang's sudden family difficulty, we recommend donating 5,600 yuan to help." After indicating their final donation, participants completed an attention check.

We measured evaluation apprehension using Spencer et al.'s (1999) 4-item scale (e.g., "If I perform poorly, Manager Liu/Algorithm CompNet will look down on me"; $\alpha = 0.90$). We also measured mind perception using Bigman and Gray's (2018) 6-item scale assessing agency (communication, thinking, planning; $\alpha = 0.76$) and experience (feeling pain, pleasure, fear; $\alpha = 0.94$). The overall scale showed $\alpha = 0.87$. Finally, participants reported AI familiarity, AI knowledge, gender, and age.

5.2.1 Effect of Supervisor Type on Compliance

Independent samples t-test showed significantly higher compliance for human supervisors ($M = 1.02$, $SD = 1.10$) than AI supervisors ($M = 0.67$, $SD = 0.89$), $t(248) = 2.75$, $p = 0.006$, Cohen's $d = 0.35$. ANCOVA controlling for gender, age, AI familiarity, and AI knowledge maintained significance, $F(1, 244) = 6.73$, $p = 0.010$, $\eta^2_p = 0.027$.

5.2.2 Mediation Analysis for Evaluation Apprehension

Zero-order correlation revealed a significant positive relationship between evaluation apprehension and compliance ($r = 0.165$, $p = 0.009$). Independent samples t-test showed human supervisors elicited higher evaluation apprehension ($M = 4.25$, $SD = 1.49$) than AI supervisors ($M = 3.83$, $SD = 1.56$), $t(248) = 2.21$, $p = 0.028$, Cohen's $d = 0.28$.

Using Hayes's (2013) PROCESS Model 4 with 5,000 bootstrap samples, evaluation apprehension showed a significant indirect effect (effect = -0.04 , 95% CI $[-0.124, -0.003]$). The direct effect remained significant (effect = -0.31 , 95% CI $[-0.559, -0.059]$), indicating partial mediation [Figure 2: see original paper].

5.2.3 Mediation Analysis for Mind Perception

Zero-order correlations showed significant positive relationships between overall mind perception ($r = 0.163$, $p = 0.010$), agency ($r = 0.131$, $p = 0.038$), and experience ($r = 0.153$, $p = 0.015$) with compliance. Human supervisors were perceived as having greater overall mind ($M = 5.87$ vs. 4.24 , $t(248) = 13.76$, $p < 0.001$, $d = 1.74$), agency ($M = 5.98$ vs. 5.41 , $t(248) = 5.10$, $p < 0.001$, $d = 0.65$), and experience ($M = 5.75$ vs. 3.06 , $t(248) = 16.88$, $p < 0.001$, $d = 2.44$).

PROCESS Model 4 revealed that overall mind perception did not mediate the effect (95% CI $[-0.332, 0.021]$). However, agency showed a significant indirect effect (effect = -0.06 , 95% CI $[-0.156, -0.007]$), with a remaining direct effect (effect = -0.29 , 95% CI $[-0.548, -0.029]$), indicating partial mediation. Experience did not mediate (95% CI $[-0.302, 0.112]$). Among specific agency dimensions, only communication ability showed significant mediation (effect = -0.05 , 95% CI $[-0.143, -0.009]$).

5.3 Discussion

Experiment 2 replicated the compliance difference across a new scenario and confirmed evaluation apprehension as a mediator. Human supervisors elicited greater evaluation apprehension, which increased compliance. While agency (particularly communication ability) also mediated, this overlaps with the interpersonal dimension of evaluation apprehension, providing convergent validation. Experiment 3 directly manipulates evaluation apprehension to establish causality.

Experiment 3: Manipulating Evaluation Apprehension

Experiment 3 used a disaster relief donation scenario to further test whether evaluation apprehension causally produces compliance differences between human and AI supervisors.

6.1.1 Participants

G*Power 3.1 indicated 265 participants were needed for a 2×2 ANOVA with small effect size ($f = 0.20$), $\alpha = 0.05$, and 90% power. We recruited 272 valid participants ($M_{age} = 30.74 \pm 7.64$) via Credamo after excluding failed attention checks.

6.1.2 Design and Procedure

This 2 (supervisor: human vs. AI) $\times 2$ (evaluation apprehension: high vs. low) between-subjects experiment randomly assigned participants to four conditions. Participants imagined being a technical employee who received a 10,000 yuan bonus during a company disaster relief campaign for a 6.2-magnitude earthquake. They initially indicated their donation amount (0-10,000 yuan).

Management information differed from previous experiments by including evaluation apprehension manipulation. High evaluation apprehension participants read: “Manager Chen (or Algorithm LR500) will evaluate all your words and actions, including this donation, creating a personal evaluation report.” Low evaluation apprehension participants read: “Manager Chen (or Algorithm LR500) will evaluate only your performance, excluding this donation.” All participants then received advice: “When disaster strikes, help comes from all sides. We recommend donating 5,600 yuan.” After indicating final donations and completing an attention check, participants completed the evaluation apprehension scale ($\alpha = 0.93$) as a manipulation check and reported demographics.

6.2.1 Manipulation Check

Independent samples t-test confirmed the manipulation: high evaluation apprehension participants reported greater concern ($M = 4.63$, $SD = 1.42$) than low evaluation apprehension participants ($M = 2.89$, $SD = 1.32$), $t(270) = -10.46$, $p < 0.001$, Cohen’s $d = 1.27$.

6.2.2 Interaction Effect

ANOVA revealed a significant main effect of evaluation apprehension, $F(1, 268) = 24.12$, $p < 0.001$, $\eta^2_p = 0.08$, with higher compliance under high ($M = 0.61$, $SD = 0.52$) versus low ($M = 0.26$, $SD = 0.65$) evaluation apprehension. The supervisor main effect was non-significant ($p = 0.384$), but the interaction was significant, $F(1, 268) = 4.59$, $p = 0.033$, $\eta^2_p = 0.02$ [Figure 3: see original paper].

Simple effects analysis showed that under high evaluation apprehension, AI supervisor compliance ($M = 0.50$, $SD = 0.47$) was significantly lower than human supervisor compliance ($M = 0.72$, $SD = 0.55$), $F(1, 268) = 4.54$, $p = 0.034$. Under low evaluation apprehension, no significant difference emerged, $F(1, 268) = 0.81$, $p = 0.370$.

6.3 Discussion

By manipulating evaluation apprehension, Experiment 3 confirmed it as a key mechanism. Only when evaluation apprehension was high did people preferentially comply with human versus AI supervisors' moral advice. When evaluation apprehension was low, supervisor type effects disappeared. This establishes evaluation apprehension as a causal mechanism. Experiment 4 now explores the boundary condition of anthropomorphic tendency.

Experiment 4: Moderating Effect of Individual Anthropomorphic Tendency

Experiment 4 examined whether individual differences in anthropomorphic tendency moderate the supervisor effect, predicting that supervisor type would influence compliance only among those low in anthropomorphism. The scenario involved charitable donations for a colleague's ill child.

7.1.1 Participants

Monte Carlo simulations indicated 250 participants were needed. We recruited 269 valid participants (180 female; $M_{age} = 30.21$, $SD = 7.59$) via Credamo, with 134 in the human condition and 135 in the AI condition.

7.1.2 Design and Procedure

This single-factor between-subjects experiment randomly assigned participants to human or AI supervisor conditions. Participants imagined being an employee who received a 10,000 yuan bonus while a colleague's 5-year-old child needed surgery for congenital heart disease. They initially indicated their donation amount (0-10,000 yuan).

After management information (human expert vs. AI algorithm) and organizational charts, participants received advice: "Given the child's urgent need and the family's financial difficulty, we recommend donating 5,600 yuan." After indicating final donations and completing an attention check, participants completed the evaluation apprehension scale ($\alpha = 0.93$) and the Individual Differences in Anthropomorphism Questionnaire (IDAQ; 15 items, $\alpha = 0.92$). Finally, they reported demographics.

7.2.1 Effect of Supervisor Type on Compliance

Independent samples t-test showed significantly higher compliance for human supervisors ($M = 0.72$, $SD = 0.72$) than AI supervisors ($M = 0.47$, $SD = 0.63$), $t(267) = 2.99$, $p = 0.003$, Cohen's $d = 0.37$.

7.2.2 Mediation of Evaluation Apprehension

PROCESS Model 4 revealed a significant indirect effect of evaluation apprehension (effect = -0.05 , 95% CI $[-0.089, -0.026]$). The direct effect became non-significant (effect = -0.07 , 95% CI $[-0.151, 0.010]$), indicating full mediation.

7.2.3 Moderating Effect of Anthropomorphic Tendency

Independent samples t-test confirmed no significant difference in anthropomorphic tendency between human ($M = 4.80$, $SD = 1.72$) and AI ($M = 4.91$, $SD = 1.85$) conditions, $t(267) = 0.51$, $p = 0.612$, ensuring group homogeneity.

PROCESS Model 1 showed that when anthropomorphic tendency was low, supervisor type significantly predicted compliance (effect = -0.30 , 95% CI $[-0.529, -0.072]$). When anthropomorphic tendency was high, this effect became non-significant (effect = -0.19 , 95% CI $[-0.424, 0.035]$) [Figure 4: see original paper].

Simple slopes analysis confirmed that low anthropomorphic tendency participants showed significantly lower compliance with AI versus human supervisors ($b = -0.30$, $SE = 0.12$, $t = -2.59$, $p = 0.010$), while high anthropomorphic tendency participants showed no significant difference ($b = -0.19$, $SE = 0.12$, $t = -1.67$, $p = 0.096$).

7.3 Discussion

Experiment 4 established anthropomorphic tendency as a boundary condition. Among those low in anthropomorphism, AI supervisor compliance was significantly lower than human supervisor compliance; this difference disappeared among those high in anthropomorphism. The experiment also replicated evaluation apprehension mediation. Experiment 5 directly manipulates AI anthropomorphism to further validate this moderation.

Experiment 5: Moderating Effect of AI Anthropomorphism

Experiment 5 directly manipulated AI anthropomorphism through textual description, comparing compliance with human supervisors, anthropomorphized AI supervisors, and non-anthropomorphized AI supervisors in a child welfare volunteer scenario.

8.1.1 Participants

Based on previous effect sizes ($d = 0.30-0.37$), G*Power indicated 321 participants were needed for a one-way three-level design with $f = 0.20$, $\alpha = 0.05$, and 90% power. We recruited 330 valid participants (198 female; Mage = 29.94, SD = 7.96) via Credamo, with 110 participants per condition.

8.1.2 Design and Procedure

This single-factor between-subjects experiment randomly assigned participants to human, anthropomorphized AI, or non-anthropomorphized AI conditions. The procedure mirrored Experiment 1a, with participants imagining volunteer work at a children's welfare home and initially indicating planned volunteer hours (0-832).

Management information was identical across conditions except for supervisor identity. Anthropomorphized AI participants read about "Archie," introduced with first-person language: "Hi! I'm Archie. I'm your department manager. I have strong analytical abilities..." Non-anthropomorphized AI participants read about "CyBer," described in third-person: "Algorithm CyBer is the department manager. It has strong analytical abilities..." A manipulation check asked: "To what extent does Archie/CyBer remind you of human characteristics?" (1 = not at all to 7 = very much).

All participants then received identical advice: "To improve child welfare, we recommend volunteering 416 hours annually." After indicating final commitments and completing an attention check, participants reported demographics.

8.2.1 Manipulation Check

Six participants whose initial plans matched the recommendation were excluded, leaving 109 in the human group, 107 in the anthropomorphized AI group, and 108 in the non-anthropomorphized AI group. Independent samples t-test confirmed that anthropomorphized AI ($M = 4.90$, $SD = 1.17$) was perceived as more human-like than non-anthropomorphized AI ($M = 4.27$, $SD = 1.39$), $t(213) = 3.59$, $p < 0.001$, Cohen's $d = 0.49$.

8.2.2 Effect of AI Anthropomorphism

One-way ANOVA revealed a significant main effect, $F(2, 321) = 8.52$, $p < 0.001$, $\eta^2_p = 0.05$. Planned contrasts showed non-anthropomorphized AI compliance ($M = 0.42$, $SD = 0.46$) was significantly lower than both anthropomorphized AI ($M = 0.62$, $SD = 0.65$) and human supervisor compliance ($M = 0.72$, $SD = 0.50$), $p_s < 0.05$ [Figure 5: see original paper].

8.3 Discussion

Experiment 5 confirmed that AI anthropomorphism moderates compliance. Enhancing AI anthropomorphism significantly increased adherence to moral behavior advice.

9. General Discussion

This research examined differences in compliance with moral behavior advice from human versus AI supervisors, investigating underlying mechanisms and boundary conditions. Through six experiments and three pilot studies ($N = 1642$, including 179 pilot participants), we found lower compliance with AI supervisors (Experiments 1a-5), with evaluation apprehension as the mediating mechanism (Experiments 2-5) and anthropomorphism as a moderator (Experiments 4-5). Diverse scenarios and samples ensured robustness.

9.1 Theoretical Contributions

First, this research reveals limitations of AI supervisors in moral guidance roles. While AI can enhance belongingness and task performance, our findings show AI supervisors are less effective at promoting moral behaviors essential for organizational culture. This extends AI leadership research by demonstrating that AI's "double-edged sword" includes deficiencies in moral advocacy. Unlike previous work focusing on AI's role in unethical behavior, we examine positive moral advice, revealing asymmetric reactions across moral valences. Our interaction-based explanation via evaluation apprehension complements attribute-based explanations like mind perception.

Second, we extend advice adoption research by incorporating AI supervisors. While consistent with prior findings of algorithm aversion, our results show unusually high compliance with human supervisors (exceeding 50%), likely because supervisors hold legitimate authority. This highlights the importance of hierarchical relationships in advice compliance.

Finally, we extend algorithm aversion research in moral domains. People reject AI involvement in moral decisions regardless of valence, viewing AI as lacking moral standing equivalent to humans. This impedes the development of ethical AI.

9.2 Practical Implications

For value alignment—ensuring AI reflects human values—our findings indicate AI currently falls short as a moral exemplar. Organizations should carefully limit AI supervisor authority in moral domains to avoid undermining ethical culture. However, anthropomorphism offers a practical solution. Our results

show that simple anthropomorphic design (human-like names, first-person communication) significantly enhances compliance. This aligns with meta-analytic evidence that robot anthropomorphism improves interaction attitudes. Companies like Xiaomi (“Xiaoai”) and Amazon (Alexa) demonstrate this approach’s feasibility. Organizations should match AI anthropomorphism levels with employees’ anthropomorphic tendencies when deploying AI management.

9.3 Limitations and Future Directions

First, our dependent variable measured only quantifiable behavioral intentions (advice weight), not qualitative choices between different moral options (e.g., sacred vs. secular values). Future research should examine more complex scenarios involving taboo trade-offs.

Second, other mechanisms may explain low AI compliance, such as perceived intentionality, motivation, or inferred benevolence. Future work should examine these variables to deepen understanding.

Third, anthropomorphism effects may be more complex. We examined only surface-level anthropomorphism, not deep anthropomorphism that might trigger uncanny valley effects. Longitudinal studies should examine whether anthropomorphism advantages persist over time. Additionally, other approaches like hybrid intelligence models or AI quotient (AIQ) may influence AI management effectiveness.

Finally, ecological validity could be improved. Our scenario experiments measured behavioral intentions, which may not translate to actual behavior due to the intention-behavior gap and social desirability biases. Future research should employ field experiments, laboratory studies, or experience sampling methods to observe actual compliance behavior in real interactions.

10. Conclusion

This research concludes: (1) People comply less with moral behavior advice from AI versus human supervisors; (2) This occurs because AI supervisors elicit lower evaluation apprehension; (3) Compliance increases when individuals have stronger anthropomorphic tendencies or when AI supervisors are more anthropomorphized.

References

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, 17(4), 351–371.

- Aguinis, H., & Glavas, A. (2012). What we know and don't know about corporate social responsibility: a review and research agenda. *Journal of Management*, 38(4), 932–968.
- Andrews, D., Bonta, J., & Wormith, J. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency*, 52(1), 7–27.
- Arango, L., Singaraju, P. S., & Niinen, O. (2023). Consumer responses to AI-generated charitable giving ads. *Journal of Advertising*, 52(4), 486–530.
- Bailey, P. E., Leon, T., Elnor, N. C., Moustafa, A. A., & Weidemann, G. (2023). A meta-analysis of the weight of advice in decision-making. *Current Psychology*, 42, 24516–24541.
- Banach, M., Lewek, J., Surma, S., Person, P. E., Sahebkar, A., Martin, S. S., Bajraktari, G., Henein, M. Y., Reiner, Ž., Bielecka-Dąbrowa, A., & Bytyçi, I. (2023). The association between daily step count and all-cause and cardiovascular mortality: A meta-analysis. *European Journal of Preventive Cardiology*, 30(18), 1975–1985.
- Barclay, P., & Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society B: Biological Sciences*, 274(1610), 749–753.
- Barnes, C. M., Lucianetti, L., Bhave, D. P., & Christian, M. S. (2015). “You wouldn't like me when I'm sleepy”: Leaders' sleep, daily abusive supervision, and work unit engagement. *Academy of Management Journal*, 58(5), 1419–1437.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370.
- Beijer, S., Peccei, R., van Veldhoven, M., & Paauwe, J. (2021). The turn to employees in the measurement of human resource practices: A critical review and proposed way forward. *Human Resource Management Journal*, 31(1), 1–17.
- Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, 181, 21–34.
- Bigman, Y. E., Wilson, D., Arnestad, M. N., Waytz, A., & Gray, K. (2023). Algorithmic discrimination causes less moral outrage than human discrimination. *Journal of Experimental Psychology: General*, 152(1), 4–27.
- Bigman, Y., Waytz, A., Alterovitz, R., & Gray, K. (2019). Holding robots responsible: The elements of machine morality. *Trends in Cognitive Sciences*, 23(5), 365–368.
- Bjugstad, K., Thach, E. C., Thompson, K. J., & Morris, A. A. (2006). A fresh look at followership: A model for matching followership and leadership styles. *Journal of Behavioral & Applied Management*, 7(3), 304–319.

- Blair, A., & Saffidine, A. (2019). AI surpasses humans at six-player poker. *Science*, *365*(6456), 864–865.
- Bolino, M. C., & Grant, A. M. (2016). The bright side of being prosocial at work, and the dark side, too: A review and agenda for research on other-oriented motives, behavior, and impact in organizations. *The Academy of Management Annals*, *10*(1), 599–670.
- Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, *101*, 127–151.
- Bonezzi, A., & Ostinelli, M. (2021). Can algorithms legitimize discrimination? *Journal of Experimental Psychology: Applied*, *27*(2), 447–459.
- Bonnefon, J., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573–1576.
- Bordia, P., Irmer, B. E., & Abusah, D. (2006). Differences in sharing knowledge interpersonally and via databases: The role of evaluation apprehension and perceived benefits. *European Journal of Work and Organizational Psychology*, *15*(3), 262–280.
- Bostrom, N., & Yudkowsky, E. (2011). The ethics of Artificial Intelligence. In K. Frankish (Ed.). *Cambridge handbook of artificial intelligence*. Cambridge: Cambridge University Press.
- Broadbent, E. (2017). Interaction with robots: The truths we reveal about ourselves. *Annual Review of Psychology*, *68*, 627–652.
- Broom, D. M. (2006). The evolution of morality. *Applied Animal Behaviour Science*, *100*(1-2), 20–28.
- Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications: Profound change is coming, but roles for humans remain. *Science*, *358*(6370), 1530–1534.
- Bucher, E., Fieseler, C., & Lutz, C. (2019). Mattering in digital labor. *Journal of Managerial Psychology*, *34*(4), 307–324.
- Burger, J. M., Messian, N., Patel, S., del Prado, A., & Anderson, C. (2004). What a coincidence! The effects of incidental similarity on compliance. *Personality and Social Psychology Bulletin*, *30*(1), 35–43.
- Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239.
- Chan, I. C. C., Lam, L. W., Chow, C. W., Fong, L. H. N., & Law, R. (2017). The effect of online reviews on hotel booking intention: The role of reader-reviewer similarity. *International Journal of Hospitality Management*, *66*, 54–65.

- Chen, F., Sengupta, J., & Zheng, J. (2023). When products come alive: Interpersonal communication norms induce positive word of mouth for anthropomorphized products. *Journal of Consumer Research*, 49(6), 1032–1052.
- Cram, W. A., Wiener, M., Tarafdar, M., & Benlian, A. (2022). Examining the impact of algorithmic control on Uber drivers' technostress. *Journal of Management Information Systems*, 39(2), 426–453.
- de Cremer, D. (2017). CC'ing the boss on email makes employees feel less trusted. *Harvard Business Review*. Retrieved January 2024, <https://hbr.org/2017/04/ccing-the-boss-on-email-makes-employees-feel-less-trusted>
- de Cremer, D. (2020). *Leadership by algorithm: Who leads and who follows in the AI era?* Basingstoke, Hampshire: Harriman House.
- de Freitas, J., Agarwal, S., Schmitt, B., Haslam, N. (2023). Psychological factors underlying attitudes towards AI tools. *Nature Human Behaviour*, 7, 1845–1854.
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668.
- Dietvorst, B. J., & Bartels, D. M. (2022). Consumers object to algorithms making morally relevant tradeoffs because of algorithms' consequentialist decision strategies. *Journal of Consumer Psychology*, 32(3), 406–424.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126.
- Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and App-work in the gig economy: A research agenda for employment relations and HRM. *Human Resource Management Journal*, 30(1), 114–132.
- Ellemers, N., Toorn, J. V. D., Paunov, Y., & Leeuwen, T. V. (2019). The psychology of morality: A review and analysis of empirical studies published from 1940 through 2017. *Personality and Social Psychology Review*, 23(4), 332–366.
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886.
- Faraji-Rad, A., Samuelsen, B. M., & Warlop, L. (2015). On the persuasiveness of similar others: The role of mentalizing and the feeling of certainty. *Journal of Consumer Research*, 42(3), 458–471.
- Fast, N. J., & Schroeder, J. (2020). Power and decision making: New directions for research in the age of artificial intelligence. *Current Opinion in Psychology*, 33, 172–176.

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A-G. (2009). Statistical power analyses using *GPower 3.1: Test for correlation and regression analyses*. *Behavior Research Methods*, 41*(4), 1149–1160.
- Feinberg, M., Kovacheff, C., Teper, R., & Inbar, Y. (2019). Understanding the process of moralization: How eating meat becomes a moral issue. *Journal of Personality and Social Psychology*, 117(1), 50–72.
- 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.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting & Social Change*, 114(1), 254–280.
- Fronczek, L. P., Mende, M., Scott, M. L. Nenkov, G. Y., & Gustafsson, A. (2017). Friend or foe? Can anthropomorphizing self-tracking devices backfire on marketers and consumers? *Journal of the Academy of Marketing Science*, 51(5), 1075–1097.
- Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines*, 30(3), 411–437.
- Garland, H. & Brown, B. R. (1972). Face-saving as affected by subjects' sex, audiences' sex and audience expertise. *Sociometry*, 35(2), 280–289.
- Garvey, A. M., Kim, T., & Duhachek, A. (2023). Bad news? Send an AI. Good news? Send a human. *Journal of Marketing*, 87(1), 10–25.
- Geerts, J., de Wit, J., & de Rooji, A. (2021). Brainstorming with a social robot facilitator: Better than human facilitation due to reduced evaluation apprehension? *Frontiers in Robotics and AI*, 8, 657291.
- Gino, F., Brooks, A. W., & Schweitzer, M. E. (2012). Anxiety, advice, and the ability to discern: Feeling anxious motivates individuals to seek and use advice. *Journal of Personality and Social Psychology*, 102(3), 497–512.
- Gino, F., Shang, J., & Croson, R. (2009). The impact of information from similar or different advisors on judgments. *Organizational Behavior and Human Decision Processes*, 108, 287–302.
- Gladden, M. E. (2014). The social robot as “charismatic leader”: A phenomenology of human submission to nonhuman power. *Frontiers in Artificial Intelligence and Applications*, 273, 329–339.
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660.
- Grag, S., Sinha, S., Kar, A., & Mani, M. (2022). A review of machine learning applications in human resource management. *International Journal of Productivity and Performance Management*, 71(5), 1590–1610.

- Gratch, J., & Fast, N. J. (2022). The power to harm: AI assistants pave the way to unethical behavior. *Current Opinion in Psychology*, *47*, 101382.
- Gray, H., Gray, K., & Wegner, D. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619.
- Gray, K., & Wegner, D. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125–130.
- Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, *96*(3), 505–520.
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101–124.
- Guzman, A. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication*, *1*, 37–54.
- Hagendorff, T. (2024). Deception abilities emerged in large language models. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(24), e2317967121.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Random House.
- Halbusi, H. A., Ruiz-Palomino, P., Morales-Sánchez, R., & Fattah, F. A. M. A. (2021). Managerial ethical leadership, ethical climate and employee ethical behavior: Does moral attentiveness matter? *Ethics & Behavior*, *31*(8), 604–627.
- Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology*, *91*(3), 436–455.
- Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*, 117–133.
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.
- Hertz, N., & Wiese, E. (2019). Good advice is beyond all price, but what if it comes from a machine? *Journal of Experimental Psychology: Applied*, *25*(3), 386–395.
- Hertz, S. G., & Krettenauer, T. (2016). Does moral identity effectively predict moral behavior: A meta-analysis. *Review of General Psychology*, *20*(2), 129–140.
- Hoch, J. E., Bommer, W. H., Dulebohn, J. H., & Wu, D. (2018). Do ethical, authentic, and servant leadership explain variance above and beyond transformational leadership? A meta-analysis. *Journal of Management*, *44*(2), 501–529.

- Höddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, *116*, 106635.
- Holford, W. (2022). An ethical inquiry of the effect of cockpit automation on the responsibilities of airline pilots: Dissonance or meaningful control? *Journal of Business Ethics*, *176*(1), 141–157.
- Holthöwer, J., & van Doorn, J. (2023). Robots do not judge: Service robots can alleviate embarrassment in service encounters. *Journal of the Academy of Marketing Science*, *51*, 767–784.
- Hu, X., Li, M., Wang, D., & Yu, F. (2024). Reactions to immoral AI decisions: The moral deficit effect and its underlying mechanism. *Chinese Science Bulletin*, *69*(11), 1406–1416.
- Hur, J. D., Koo, M., & Hofmann, M. (2015). When temptations come alive: How anthropomorphism undermines self-control. *Journal of Consumer Research*, *42*(2), 340–358.
- Inesi, M. E., Adams, G. S., & Gupta, A. (2021). When it pays to be kind: The allocation of indirect reciprocity within power hierarchies. *Organizational Behavior and Human Decision Processes*, *165*, 115–126.
- Ivancevich, J. M., Konopaske, R., & Matteson, M. T. (2005). *Organizational behavior and management* (7th ed.). Boston: McGraw-Hill Irwin.
- Jabagi, N., Croteau, A. M., Audebrand, L. K., & Marsan, J. (2019). Gig-workers' motivation: Thinking beyond carrots and sticks. *Journal of Managerial Psychology*, *34*(4), 192–213.
- Jackson, J., Yam, K., Tang, P., Liu, T., & Shariff, A. (2023). Exposure to robot preachers undermines religious commitment. *Journal of Experimental Psychology: General*, *152*(12), 3344–3358.
- Jago, A. S., Raveendhran, R., Fast, N., & Gratch, J. (2024). Algorithmic management diminishes status: An unintended consequence of using machines to perform social roles. *Journal of Experimental Social Psychology*, *110*, 104553.
- Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology*, *96*(3), 521–537.
- Jarrassé, N., Sanguinetti, V., & Burdet, E. (2014). Slaves no longer: Review on role assignment for human-robot joint motor action. *Adaptive Behavior*, *22*, 70–82.
- Jebari, K., & Lundbory, J. (2021). Artificial superintelligence and its limits: Why AlphaZero cannot become a general agent. *AI & Society*, *36*, 807–815.
- Jia, N., Luo, X., Fang, Z., & Liao, C. (2023). When and how artificial intelligence augments employee creativity. *Academy of Management Journal*, *67*(1),

5–32.

Jiang, L., Hoegg, J., Dahl, D. W., & Chattopadhyay, A. (2010). The persuasive role of incidental similarity on attitudes and purchase intentions in a sales context. *Journal of Consumer Research*, *36*(5), 778–791.

Jiang, Z., & Hu, X. (2016). Knowledge sharing and life satisfaction: The roles of colleague relationships and gender. *Social Indicators Research*, *126*(1), 379–394.

Jung, D., Dorner, V., Glaser, F., & Morana, S. (2018). Robo-advisory: Digitalization and automation of financial advisory. *Business and Information Systems Engineering*, *60*(1), 81–86.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*(1), 366–410.

Kelly, S., Kaye, S., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, *77*, 101925.

Kim, H. C., & Kramer, T. (2015). Do materialists prefer the “brand-as-servant”? The interactive effects of anthropomorphized brand roles and materialism on consumer responses. *Journal of Consumer Research*, *42*(2), 284–299.

Kim, Y., Chin, C., Peng, S., Cai, H., & Tov, W. (2010). Explaining east-west differences in the likelihood of making favorable self-evaluations: The role of evaluation apprehension and directness of expression. *Journal of Cross-Cultural Psychology*, *41*(1), 62–75.

Kinowska, H., & Sienkiewicz, L. J. (2023). Influence of algorithmic management practices on workplace well-being – evidence from European organisations. *Information Technology & People*, *36*(8), 21–42.

Kish-Gephart, J. J., Harrison, D. A., & Trevino, L. K. (2010). Bad apples, bad cases, and bad barrels: Meta-analytic evidence about sources of unethical decisions at work. *Journal of Applied Psychology*, *95*(1), 1–31.

Knobe, J. (2003). Intentional action and side-effects in ordinary language. *Analysis*, *63*, 190–194.

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, *8*(3), 239–260.

Konya-Baumbach, E., Biller, M., & von Janda, S. (2023). Someone out there? A study on the social presence of anthropomorphized chatbots. *Computers in Human Behavior*, *139*, 107513.

- Kormos, C., & Gifford, R. (2014). The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *Journal of Environmental Psychology, 40*, 359–371.
- Koslov, K., Mendes, W. B., Pajtas, P. E., & Pizzagalli, D. A. (2011). Asymmetry in resting intracortical activity as a buffer to social threat. *Psychological Science, 22*(5), 641–649.
- Kuchenbrandt, D., Eyssel, F., Bobinger, S., & Neufeld, M. (2013). When a robot's group membership matters. *International Journal of Social Robotics, 5*(3), 409–417.
- Ladak, A. (2024). What would qualify an artificial intelligence for moral standing? *AI and Ethics, 4*(2), 213–228.
- Ladeira, W., Perin, M. G., & Santini, F. (2023). Acceptance of service robots: A meta-analysis in the hospitality and tourism industry. *Journal of Hospitality Marketing & Management, 32*(6), 694–716.
- Lam, C. F., Wan, W. H., & Roussin, C. J. (2015). Going the extra mile and feeling energized: An enrichment perspective of organizational citizenship behaviors. *Journal of Applied Psychology, 101*(3), 379–391.
- Lan, H., Tang, X., Ye, Y., & Zhang, H. (2024). Abstract or concrete? The effects of language style and service context on continuous usage intention for AI voice assistants. *Humanities and Social Sciences Communications, 11*, 99.
- Langer, M., & Landers, R. N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior, 123*, 106878.
- Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment, 27*(3), 217–234.
- Lanz, L., Briker, R., & Gerpott, F. H. (2024). Employees adhere more to unethical instructions from human than AI supervisors: Complementing experimental evidence with machine learning. *Journal of Business Ethics, 189*, 625–646.
- Larkin, C., Otten, C. D., & Arvai, J. (2021). Paging Dr. JARVIS! Will people accept advice from artificial intelligence for consequential risk management decisions? *Journal of Risk Research, 25*(4), 407–422.
- Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science, 52*(1), 111–127.
- Leary, M. R. (1983). A brief version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin, 9*(3), 371–375.
- Leary, M. R. (1995). *Self-presentation: Impression management and interpersonal behavior*. Madison, WI: Brown & Benchmark.

- Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, *107*(1), 34–47.
- Lecher, C. (2019). How Amazon automatically tracks and fires warehouse workers for ‘productivity’: Documents show how the company tracks and terminates workers. *The Verge*. Retrieved January 25, 2024, from <https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centers-productivity-firing-terminations>
- Lee, J., Lee, D., & Lee, J. (2024). Influence of rapport and social presence with an AI psychotherapy chatbot on users’ self-disclosure. *International Journal of Human-Computer Interaction*, *40*(7), 1620–1631.
- Lee, S., Lee, N., & Sah, Y. J. (2020). Perceiving a mind in a chatbot: Effect of mind perception and social cues on co-presence, closeness and intention to use. *International Journal of Human Computer Interaction*, *36*(10), 930–940.
- Lee, Z., & Sargeant, A. (2011). Dealing with social desirability bias: An application to charitable giving. *European Journal of Marketing*, *45*(5), 703–719.
- Lefowitz, J. (2006). The constancy of ethics amidst the changing world of work. *Human Resource Management Review*, *16*(2), 245–268.
- Lehdonvirta, V. (2018). Flexibility in the gig economy: Managing time on three online piecework platforms. *New Technology, Work, and Employment*, *33*(1), 13–29.
- Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., & Kasper, G. (2019). The challenges of algorithm-based HR decision-making for personal integrity. *Journal of Business Ethics*, *160*, 377–392.
- Lemaignan, S., Fink, J., & Dillenbourg, P. (2014, March). The dynamics of anthropomorphism in robotics. Paper presented at the meeting of Proceedings of the 2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI), New York, United States.
- Leo, X., & Huh, Y. (2020). Who gets the blame for service failures? Attribution of responsibility toward robot versus human service providers and service firms. *Computers in Human Behavior*, *113*, 106520.
- Li, X., & Sung, Y. (2021). Anthropomorphism brings us closer: The mediating role of psychological distance in User–AI assistant interactions. *Computers in Human Behavior*, *118*, 106680.
- Liu, M., Tang, X., Xia, S., Zhang, S., Zhu, Y., & Meng, Q. (2023). Algorithm aversion: Evidence from ridesharing drivers. *Management Science*. Advance online publication. <https://doi.org/10.1287/mnsc.2022.02475>
- Longoni, C., Bonezzi, A., Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629–650.

- Louie, T. A., & Obermiller, C. (2000). Gender stereotypes and social-desirability effects on charity donation. *Psychology & Marketing*, 17(2), 121–136.
- Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015, March). Sacrifice one for the good of many? People apply different moral norms to human and robot agents. Paper presented at the meeting of Proceedings of the 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Portland, Oregon, United States.
- Malle, B. F., & Scheutz, M. (2017). Moral competence in social robots. In W. Wallach, & P. Asaro (Eds), *Machine ethics and robot ethics*. Routledge.
- Maninger, T., & Shank, D. B. (2022). Perceptions of violations by artificial and human actors across moral foundations. *Computers in Human Behavior Report*, 5, 100154.
- Martin, B. A. S., Jin, H. S., Wang, D., Nguyen, H., Zhan, K., Wang, Y. X. (2020). The influence of consumer anthropomorphism on attitudes towards artificial intelligence trip advisors. *Journal of Hospitality and Tourism Management*, 44, 108–111.
- May, F., & Monga, A. (2014). When time has a will of its own, the powerless don't have the will to wait: Anthropomorphism of time can decrease patience. *Journal of Consumer Research*, 40(5), 924–942.
- McCroskey, J. C. (1977). Oral communication apprehension: A summary of recent theory and research. *Human Communication Research*, 4(1), 78–96.
- Mell, J., Lucas, G., Mozgai, S., & Gratch, J. (2020). The effects of experience on deception in human-agent negotiation. *Journal of Artificial Intelligence Research*, 68, 633–660.
- Millet, K., Buehler, F., Du, G., & Kokkoris, M. (2023). Defending humankind: Anthropocentric bias in the appreciation of AI art. *Computers in Human Behavior*, 143, 107707.
- Möhlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. W. (2021). Algorithmic management of work on online labor platforms: When matching meets control. *MIS Quarterly*, 45(4), 1999–2022.
- Mölein, F. (2018). Robots in the boardroom: Artificial intelligence and corporate law. In B. Woodrow., P. Ugo. (Eds.), *Research Handbook on the Law of Artificial Intelligence* (pp. 649–650). Cheltenham: Edward Elgar Publishing.
- Mori, M. (1970). The uncanny valley. *Energy*, 7, 33–35.
- Moritz, J. M., Pomrehn, L., Steinmetz, H., & Wehner, M. (2024, July). Reactions to algorithmic decision-making in human resource management: A meta-analysis. Paper presented at the meeting of Academy of Management Proceedings, Valhalla, New York, United States.

- Munnukka, J., Talvitie-Lamberg, K., & Maity, D. (2022). Anthropomorphism and social presence in Human-Virtual service assistant interactions: The role of dialog length and attitudes. *Computers in Human Behavior*, *135*, 107343.
- Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103.
- Newman, D., Fast, N., & Harmon, D. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*, *160*, 149–167.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, *91*(3), 328–346.
- Niehoff, B. P. & Moorman, R. H. (1993). Justice as a mediator of the relationship between methods of monitoring and organizational citizenship behavior. *Academy of Management Journal*, *36*(3), 527–556.
- Niszczoła, P., & Kaszás, D. (2020). Robo-investment aversion. *PloS One*, *15*(9), e0239277.
- Noval, L. J., & Stahl, G. K. (2017). Accounting for proscriptive and prescriptive morality in the workplace: The double-edged sword effect of mood on managerial ethical decision making. *Journal of Business Ethics*, *142*(3), 603-602.
- Oh, C., Song, J., Choi, J., Kim, S., Lee, S., & Suh, B. (2018, April). I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. Paper presented at the meeting of Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, Canada.
- Paillé, P., & Boiral, O. (2013). Pro-environmental behavior at work: Construct validity and determinants. *Journal of Environmental Psychology*, *36*, 118–128.
- Parent-Rocheleau, X., & Parker, S. K. (2022). Algorithms as work designers: How algorithmic management influences the design of jobs. *Human Resources Management Review*, *32*(3), 100838.
- Parent-Rocheleau, X., Parker, S. K., Bujold, A., & Gaudet, M-C. (2024). Creation of the algorithmic management questionnaire: A six-phase scale development process. *Human Resource Management*, *63*, 25–44.
- Park, H., Ahn, D., Hosanagar, K., & Lee, J. (2021). Human-AI interaction in human resource management: Understanding why employee resist algorithmic evaluation at workplaces and how to mitigate burdens. Paper presented at the meeting of Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan.
- Park, J., Woo, S. E., & Kim, J. (2024). Attitudes towards artificial intelligence at work: Scale development and validation. *Journal of Occupational and Organizational Psychology*. Advance online publication.

- Parry, K., Cohen, M., & Bhattacharya, S. (2016). Rise of the machines: A critical consideration of automated leadership decision making in organizations. *Group & Organization Management*, *41*, 571–594.
- Pelau, C., Dabija, D. C., & Ene, I. (2021). What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. *Computers in Human Behavior*, *122*, 106855.
- Pelletier, K. L., & Bligh, M. C. (2008). The aftermath of organizational corruption: Employee attributions and emotional reactions. *Journal of Business Ethics*, *80*(4), 823–844.
- Peng, A. C., & Kim, D. (2020). A meta-analytic test of the differential pathways linking ethical leadership to normative conduct. *Journal of Organizational Behavior*, *41*(4), 348–368.
- Pickard, M. D., & Roster, C. A. (2020). Using computer automated systems to conduct personal interviews: Does the mere presence of a human face inhibit disclosure? *Computers in Human Behavior*, *105*, 106197.
- Pitardi, V., Wirtz, J., Paluch, S., & Kunz, W. H. (2022). Service robots, agency and embarrassing service encounters. *Journal of Service Management*, *33*(2), 389–414.
- Podsakoff, N. p., Whiting, S. W., Podsakoff, P. M., & Blume, B. D. (2009). Individual- and organizational-level consequences of organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology*, *94*(1), 122–141.
- Qin, X., Lu, J. G., Chen, C., Zhou, X., Gan, Y., Li, W., & Song, L. L. (2024). Artificial intelligence quotient (AIQ). *PsyArXiv Preprints*, <https://doi.org/10.31234/osf.io/qjm3r>
- Rader, C. A., Larrick, R. P., & Soll, J. B. (2017). Advice as a form of social influence: Informational motives and the consequences for accuracy. *Social Personality Psychology Compass*, *11*, e12329.
- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, *46*(1), 192–210.
- Raveendhran, R., & Fast, N. J. (2019). Technology and social evaluation: Opportunities and challenges. In R. N. Landers (Ed.), *The Cambridge Handbook of Technology and Employee Behavior* (pp. 921–943). New York: Cambridge University Press.
- Raveendhran, R., & Fast, N. J. (2021). Human judge, algorithms nudge: The psychology of behavior tracking acceptance. *Organizational Behavior and Human Decision Processes*, *164*, 11–26.
- Raveendhran, R., Fast, N. J., & Carnevale, P. J. (2020). Virtual (freedom

from) reality: Evaluation apprehension and leaders' preference for communicating through avatars. *Computers in Human Behavior*, 111, 106415.

Rodell, J. B., Booth, J., Lynch, J., & Zipay, K. (2017). Corporate volunteering climate: Mobilizing employee passion for societal causes and inspiring future charitable action. *Academy of Management Journal*, 60(5), 1662–1681.

Roesler, E., Manzey, D., & Onnasch, L. (2021). A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics*, 6(58), eabj5425.

Rosenberg, M. J. (1965). When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology*, 1(1), 28–42.

Ruttan, R. L. & Nordgren, L. F. (2021). Instrumental use erodes sacred values. *Journal of Personality and Social Psychology*, 121(6), 1223–1240.

Ryan, R. M. & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, 57(5), 749–761.

Schlenker, B. R. & Leary, M. R. (1982). Social anxiety and self-presentation: A conceptualization model. *Psychological Bulletin*, 92(3), 641–669.

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

Schwartz, E. H. (2023). Huawei adds generative AI-powered voice assistant to smartphones, beating apple and google. *Voicebot.ai*. Retrieved January 2024, <https://voicebot.ai/2023/08/08/huawei-adds-generative-ai-powered-voice-assistant-to-smartphones-beating-apple-and-google/>

Sen, S., Du, S., & Bhattacharya, C. (2016). Corporate social responsibility: A consumer psychology perspective. *Current Opinion in Psychology*, 10, 70–75.

Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review. *European Review of Social Psychology*, 12(1), 1–36.

Siemon, B. (2023). Let the computer evaluate your idea: Evaluation apprehension in human-computer collaboration. *Behaviour & Information Technology*, 42(5), 459–477.

Singh, S., Olson, E. D., & Tsai, C. H. K. (2021). Use of service robots in an event setting: Understanding the role of social presence, eeriness, and identity threat. *Journal of Hospitality and Tourism Management*, 49, 528–537.

Smith, E. R., Šabanović, S., & Fraune, M. R. (2021). Human-robot interaction through the lens of social psychological theories of intergroup behavior. *Technology, Mind, Behavior*, 1(2), 2.

- Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes*, *62*, 159–174.
- Sniezek, J. A., & van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, *84*, 288–307.
- Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 780–805.
- Song, H., Barakova, E. I., Ham, J., & Markopoulos, P. (2024). The impact of social robots' presence and roles on children's performance in musical instrument practice. *British Journal of Educational Technology*, *55*(3), 1041–1059.
- Space and Lemon Innovations. (2023). Assistants for a car driver – the new developments. *Medium*. Retrieved January 2024, <https://spaceandlemon.medium.com/assitants-for-a-car-driver-the-new-developments-3c494ba33ca5>
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28.
- Stanley, M. L., & Kay, A. C. (2024). The consequences of heroization for exploitation. *Journal of Personality and Social Psychology: Attitudes and Social Cognition*, *126*(1), 5–25.
- Stanley, M. L., Neck, C. B., & Neck, C. P. (2023a). Loyal workers are selectively and ironically targeted for exploitation. *Journal of Experimental Social Psychology*, *106*, 104442.
- Stanley, M. L., Neck, C. P., & Neck, C. B. (2023b). The dark side of generosity: Employees with a reputation for giving are selectively targeted for exploitation. *Journal of Experimental Social Psychology*, *108*, 104503.
- Stikvoort, B., Lindahl, T., & Daw, T. M. (2016). Thou shalt not sell nature: How taboo trade-offs can make us act pro-environmentally, to clear our conscience. *Ecological Economics*, *129*, 252–259.
- Sun, J., Liden, R. C., & Ouyang, L. (2019). Are servant leaders appreciated? An investigation of how relational attributions influence employee feelings of gratitude and prosocial behaviors. *Journal of Organizational Behavior*, *40*(5), 528–540.
- Tang, P. M., Koopman, J., Elfenbein, H. A., Zhang, J. H., de Cremer, D., Li, C. H., & Chan, E. T. (2022). Using robots at work during the COVID-19 crisis evokers passion decay: Evidence from field and experimental studies. *Journal of Applied Psychology*, *71*(3), 881–911.
- Tang, P. M., Koopman, J., Mai, K. M., de Cremer, D., Zhang, J. H., Reynders, P., Ng, C. T. S., & Chen, I. H. (2023). No person is an island: Unpacking

the work and after-work consequences of interacting with artificial intelligence. *Journal of Applied Psychology*, 108(11), 1766–1789.

Tang, P. M., Koopman, J., Yam, K. C., de Cremer, D., Zhang, J. H., & Reynnders, P. (2023). The self-regulatory consequences of dependence on intelligent machines at work: Evidence from field and experimental studies. *Human Resource Management*, 62(5), 721–744.

Tangney, J. P. (1992). Situational determinants of shame and guilt in young adulthood. *Personality and Social Psychology Bulletin*, 18(2), 199–206.

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, 7(7), 320–324.

Tomprou, M., & Lee, M. K. (2022). Employment relationships in algorithmic management: A psychological contract perspective. *Computers in Human Behavior*, 126, 106997.

Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, 42(9), 1600–1631.

Treviño, L. K., Weaver, G. R., & Reynolds, S. J. (2006). Behavioral ethics in organizations: A review. *Journal of Management*, 32(6), 951–990.

Tsai, C., Marshall, J. D., Choudhury, A., Serban, A., Hou, Y. T., Jung, M. F., Dionne, S. D., & Yammarino, F. J. (2022). Human-robot collaboration: A multilevel and integrated leadership framework. *The Leadership Quarterly*, 33, 101594.

Uysal, E., Alavi, S., & Bezençon, V. (2023). Anthropomorphism in artificial intelligence: A review of empirical work across domains and insights for future research. *Artificial Intelligence in Marketing*, 20, 273–308.

van Beurden, J., van de Voorde, K., & van Veldhoven, M. (2021). The employee perspective on HR practices: A systematic literature review, integration and outlook. *The International Journal of Human Resource Management*, 32(2), 359–393.

van Boven, L., Loewenstein, G., & Dunning, D. (2005). The illusion of courage in social predictions: Underestimating the impact of fear of embarrassment on other people. *Organizational Behavior and Human Decision Processes*, 96, 130–141.

van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Peterson, J. A. (2017). Domo Arigato, Mr. Roboto: Emergence of automated social presence in organizational frontlines and customers' service experiences. *Journal of Service Research*, 20(1), 1–16.

van Zoonen, W., Sivunen, A. E., & Treem, J. W. (2024). Algorithmic management of crowdworkers: Implications for workers' identity, belonging, and meaningfulness of work. *Computers in Human Behavior*, 152, 108089.

- Veetikazhi, R., Kamalanabahan, T. J., Malhotra, P., Arora, R., & Mueller, A. (2022). Unethical employee behaviour: A review and typology. *The International Journal of Human Resource Management*, 33(10), 1976–2018.
- von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*, 4(4), 404–409.
- Wang, Y., & Chuang, Y. (2024). Artificial intelligence self-efficacy: Scale development and validation. *Education and Information Technologies*, 29, 4785–4808.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070.
- Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5(3), 219–232.
- Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52, 113–117.
- Weisman, K., Dweck, C. S., & Markman, E. M. (2017). Rethinking people's conceptions of mental life. *Proceedings of the National Academy of Sciences*, 114(43), 11374–11379.
- Wesche, J. S., & Sonderegger, A. (2019). When computers take the lead: The automation of leadership. *Computers in Human Behavior*, 101, 197–209.
- Wieland, B., de Wit, J., & de Rooij, A. (2022). Electronic brainstorming with a Chatbot partner: A good idea due to increased productivity and idea diversity. *Frontiers in Artificial Intelligence*, 5, 880673.
- Wu, M., Wang, N., & Yuen, K. (2023). Deep versus superficial anthropomorphism: Exploring their effects on human trust in shared autonomous vehicles. *Computers in Human Behavior*, 141, 107614.
- Xu, L., & Yu, F. (2020). Factors that influence robot acceptance. *Chinese Science Bulletin*, 65(6), 496–510.
- Xu, L., Mehta, R., & Dahl, D. W. (2022). Leveraging Creativity in Charity Marketing: The Impact of Engaging in Creative Activities on Subsequent Donation Behavior. *Journal of Marketing*, 86(5), 79–94.
- Xu, L., Yu, F., & Peng, K. (2022). Algorithmic discrimination causes less desire for moral punishment than human discrimination. *Acta Psychologica Sinica*, 54(9), 1076–1092.
- Xu, L., Yu, F., Wu, J., Han, T., & Zhao., L. (2017). Anthropomorphism: Antecedents and consequences. *Advances in Psychological Science*, 25(11),

1942–1954.

Yam, K., Bigman, Y. E., Tang, P., Ilies, R., de Cremer, D., & Soh, H. (2021). Robots at work: People prefer- and forgive- service robots with perceived feelings. *Journal of Applied Psychology, 106*(10), 1557–1572.

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93*, 1–13.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes, 83*, 260–281.

Yaniv, I., & Milyavksy, M. (2007). Using advice from multiple sources to revise and improve judgment. *Organizational Behavior and Human Decision Processes, 103*, 104–120.

Yaniv, I., Choshen-Hillel, S., Milyarsky, M. (2011). Receiving advice on matters of taste: Similarity, majority influence, and taste discrimination. *Organizational Behavior and Human Decision Processes, 115*, 111–120.

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

Young, A. D., & Monroe, A. E. (2019). Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas. *Journal of Experimental Social Psychology, 85*, 103870.

Yu, F., & Xu, L. (2018). How to make an ethical artificial intelligence? Answer from a psychological perspective. *Global Media Journal, 5*(4), 24–42.

Yu, H., Miao, C., Chen, Y., Fauvel, S., Li, X., & Lesser, V. (2017). Algorithmic management for improving collective productivity in crowdsourcing. *Scientific Reports, 7*(1), 12541.

Yukl, G. (2013). *Leadership in organizations* (8. ed.). Edinburgh Gate: Pearson Education.

Zhao, Y., Xu, L., Yu, F., & Jin, W. (2024). Perceived opacity leads to algorithm aversion in the workplace. *Acta Psychologica Sinica, 56*(4), 497–514.

Zhou, X., Zhai, H., Delidabieke, B., Zeng, H., Cui, Y., & Cao, X. (2019). Exposure to ideas, evaluation apprehension, and incubation intervals in collaborative idea generative. *Frontiers in Psychology, 10*, Article

Zhou, Y., Fei, Z., He, Y., & Yang, Z. (2022). How human-chatbot interaction impairs charitable giving: The role of moral judgment. *Journal of Business Ethics, 178*, 849–865.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.