PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

Evidence of a social evaluation penalty for using AI

Jessica A. Reif[a,1], Richard P. Larrick[a], and Jack B. Soll[a]

Edited by Susan Fiske, Princeton University, Jamaica, VT; received December 23, 2024; accepted April 6, 2025

Despite the rapid proliferation of AI tools, we know little about how people who use them are perceived by others. Drawing on theories of attribution and impression management, we propose that people believe they will be evaluated negatively by others for using AI tools and that this belief is justified. We examine these predictions in four preregistered experiments (N = 4,439) and find that people who use AI at work anticipate and receive negative evaluations regarding their competence and motivation. Further, we find evidence that these social evaluations affect assessments of job candidates. Our findings reveal a dilemma for people considering adopting AI tools: Although AI can enhance productivity, its use carries social costs.

attribution | social evaluation | AI | impression management

Significance

As AI tools become increasingly prevalent in workplaces, understanding the social dynamics of AI adoption is crucial. Through four experiments with over 4,400 participants, we reveal a social penalty for AI use: Individuals who use AI tools face negative judgments about their competence and motivation from others. These judgments manifest as both anticipated and actual social penalties, creating a paradox where productivity-enhancing AI tools can simultaneously improve performance and damage one's professional reputation. Our findings identify a potential barrier to AI adoption and highlight how social perceptions may reduce the acceptance of helpful technologies in the workplace.

A longstanding question in social psychology is how perceivers interpret and explain others' behavior when there are multiple possible causes (1–4). For example, when someone accepts assistance completing a task, observers may conclude that this choice reflects dispositional qualities (e.g., a lack of ability or motivation to complete the task without help) or situational factors (e.g., receiving assistance is the norm for the task). Attribution theory posits that observers tend to favor dispositional explanations relative to situational ones (5, 6), suggesting they may attribute the use of assistance to personal deficits rather than circumstantial causes. Professional image concerns may therefore motivate people to avoid appearing dependent on external assistance (7, 8).

The attribution processes that shape evaluations of help-seeking behavior can be fruitfully extended to understand the social implications of AI use in today's workplaces. AI technologies present a dilemma to the people who use them. On the one hand, AI can enhance human performance on a variety of tasks (9, 10). People thus have strong incentives to use AI, as it might improve their performance at work. On the other hand, AI represents a powerful form of assistance. Consequently, using AI may raise doubts about one's own abilities and motivation. Consistent with this notion, a recent industry survey found that apprehension about being perceived as lazy ranks among the top concerns of people who use AI at work (11). Further, numerous reports suggest that people actively *conceal* their AI use in professional settings (12, 13). This apparent tension between AI's documented benefits and people's reluctance to use it raises a critical question: are people who use AI actually evaluated less favorably than people who receive other forms of assistance at work? Extending theories of attribution, we propose that observers will be likely to make (negative) dispositional inferences about people who receive help from AI relative to people who receive other forms of help.

In four preregistered studies, we examine this prediction from the lens of both the help recipient and observer. In *Study* 1, we show that people who receive help from AI believe they will be evaluated as lazier, less competent, and less diligent than people who receive similar help from non-AI technologies. In *Study* 2, we demonstrate that this fear is justified: observers perceive people who get help from AI as lazier, less competent, and less diligent than people who get help from other sources. *Study* 3 shows that managers who do not use AI themselves may act on their negative assumptions of people who use AI in an incentive-compatible hiring task. Finally, *Study* 4 shows that perceptions of laziness mediate the relationship between AI use and assessments of poor task fit in a hiring scenario.

Attribution Theory and Emerging Technologies

The notion that using technologies that reduce the need for effort or ability can cast doubt on one's competence and motivation has echoed in debates over new tools for centuries. For example, Plato's *Phaedrus* (370 BC) recounts a question about whether people who relied on a new invention for learning (writing) would ever develop true wisdom (14). More recently, educators have questioned how using tools such as calculators would affect students' ability to develop mathematics skills, and studies have documented patients' tendency to assume that physicians who use diagnostic aids are less capable (15, 16).

Author affiliations: [a]Fuqua School of Business, Management & Organizations, Duke University, Durham, NC 27701

Author contributions: J.A.R., R.P.L., and J.B.S. designed research; J.A.R. performed research; J.A.R. and J.B.S. analyzed data; and J.A.R., R.P.L., and J.B.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: jessica.reif@duke.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2426766122/-/DCSupplemental.

Published May 8, 2025.

Unlike previous tools that simply performed specific operations or made predictions, AI tools may be perceived as more agentic because they can learn from experience and operate more autonomously (17). Such powerful tools may *intensify* doubts related to the ability and effort of their operators.

Effort and ability have long been among the primary metrics by which people are evaluated in professional environments (18, 19), and thus people aim to project these valued qualities to others (7, 20). Assistance of any kind creates attributional ambiguity, raising impression management concerns for recipients. When a person receives help to perform a task, observers must determine how much credit is due to the person versus the assistance (21). In doing so, they may imagine counterfactual scenarios in which the person did not receive the assistance and anticipate what outcome might have occurred (22). This process extends to judgments about effort, where observers might consider both the actual effort they observed and counterfactual possibilities about how the outcome might have been different had the target exerted more effort (23). Receiving help may thus cast a shadow of doubt over one's ability and willingness to exert effort (3, 24). Conscious of this ambiguity, recipients of help may worry that observers will discount their competence and motivation and withhold any indication that they received help (8, 25). These social evaluation concerns may be justified, as observers could interpret the decision to utilize assistance as a signal that the recipient is not willing or able to perform the task themselves. Behaviors that are not considered mainstream—or part of the consensus—are especially likely to elicit dispositional attributions (24). Emerging technologies are, by definition, new, and therefore, their use is unlikely to be perceived as customary. Extending attribution theory, we propose that the use of emerging technologies that reduce the need for effort or ability is especially likely to evoke negative dispositional inferences about their operators.

The anticipation of these negative perceptions presents a dilemma to people considering whether to adopt AI: using AI may simultaneously enhance their productivity but undermine others' perceptions of their competence and motivation. Although there is a significant body of work examining how people perceive AI systems themselves (26), we know little about how evaluators perceive the people who use them. Understanding whether receiving help from AI in fact leads to a social evaluation penalty is crucial for anticipating and addressing challenges related to the adoption of AI.

Results

Study 1. We first examined whether employees would be more reluctant to disclose the use of an AI tool at work relative to another (non-AI) tool and how they expected to be perceived for using each tool. We recruited a sample of 500 online participants and asked them to imagine that they recently started using either a "generative AI tool" *(AI Tool* condition) or a "dashboard creation tool" (*Non-AI Tool* condition) to perform a task at work. Participants then rated the likelihood that they would disclose their use of this tool to their manager and colleagues as well as how they *expected* to be perceived by others on four dimensions: laziness, replaceability, competence, and diligence. Because AI tools may be perceived as more agentic than non-AI tools (and thus more worthy of receiving credit for successful task completion), we predicted that AI tool users would believe others would view them as lazier, more replaceable, less competent, and less diligent that non-AI tool users. We also expected that they would be less likely to disclose the use of these AI tools to managers and colleagues.

We conducted *t* tests to examine differences between the *AI tool* and *Non-AI Tool* conditions for each of the six dependent variables measured in this study. Fig. 1 depicts the effect sizes (Cohen's *d*) for all six *t* tests. First, we examined whether participants in the *AI Tool* condition believed they would be evaluated as lazier and more replaceable than participants in the *Non-AI Tool* condition. Consistent with our hypothesis, participants in the *AI Tool* condition believed they would be perceived as lazier (M = 3.25, SD = 1.49) than participants in the *Non-AI Tool* condition (M = 2.72, SD = 1.41), $t(492.8) = 4.08$, 95% CI [0.28, 0.79], $P < 0.001$). They also reported that they would be perceived as more replaceable in the *AI Tool* condition (M = 3.39, SD = 1.67) than in the *Non-AI* Tool condition (M = 2.83, SD = 1.63), $t(494.3) = 3.84$, 95% CI [0.28, 0.86], $P < 0.001$).

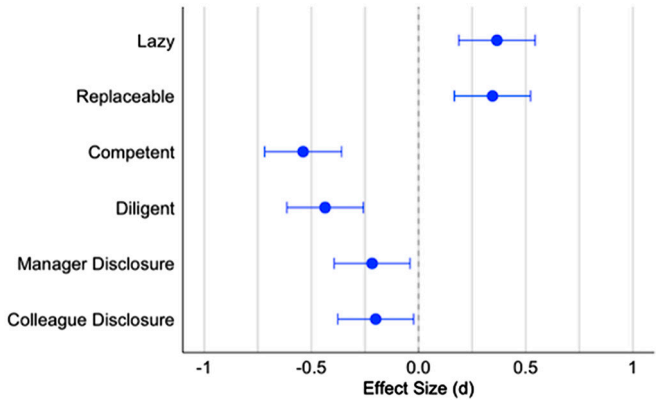

Fig. 1. Effect sizes for differences in expected perceptions and disclosure to others (*Study* 1). Note: Positive d values indicate higher values in the AI Tool condition, while negative d values indicate lower values in the AI Tool condition. N = 497. Error bars represent 95% CI. Correlations among variables range from | r |= 0.53 to 0.88.

We next examined how participants believed others would evaluate them on the two dimensions of agency we measured in this study: competence and diligence. Participants reported that they believed others would judge them as less competent in the *AI Tool* condition (M = 4.72, SD = 1.46) than in the *Non-AI Tool* condition (M = 5.45, SD = 1.22), $t(477.5) = 6.00$, 95% CI [–0.96, –0.49], $P < 0.001$). Similarly, they reported that they expected to be perceived as less diligent in the *AI Tool* condition (M = 4.66, SD = 1.44) than in the *Non-AI Tool* condition (M = 5.25, SD = 1.28), $t(487.5) = 4.86$, 95% CI [–0.83, –0.35], $P < 0.001$). These results support the notion that people known to use more agentic technologies believe they may be evaluated as less agentic themselves.

Finally, we examined the two variables related to disclosure. Participants in the *AI Tool* condition reported that they would be less likely to disclose the use of the tool to their managers (M = 4.91, SD = 1.59) than participants in the *Non-AI Tool* condition (M = 5.25, SD = 1.55), $t(494.3) = 2.42$, 95% CI [–0.62, –0.06], $P = 0.016$). Participants in the *AI Tool* condition also reported less willingness to disclose the use of the tool to colleagues (M = 4.85, SD = 1.58) than participants in the *Non-AI Tool* condition (M = 5.17, SD = 1.57), $t(494.8) = 2.23$, 95% CI [–0.59, –0.04], $P = 0.026$). These results are consistent with our prediction that people who use AI tools may be reluctant to disclose their use to others at work.

Study 2. We next tested whether evaluators assess employees known to receive help from AI more negatively than people who receive other forms of help or people who received no help at all. We recruited 1,215 online participants to complete a study in which they read a paragraph about an employee and rated how lazy they perceive the employee to be, as well as how they view the employee on six dimensions of agency: competent, diligent, ambitious, independent, self-assured, and dominant (27). The paragraph presented was randomly selected from 384 unique stimuli that systematically manipulated gender, occupation,

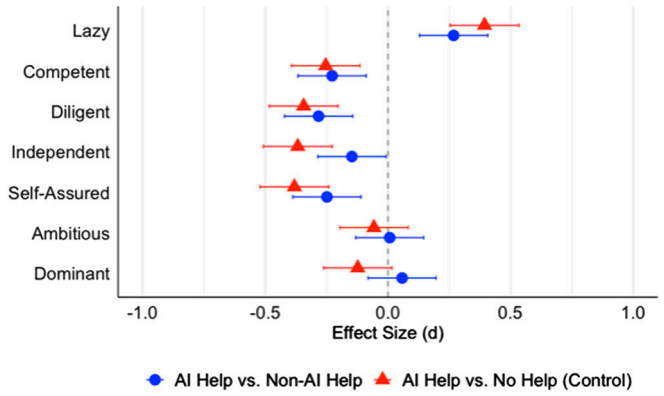

Fig. 2. Differences in evaluations for AI help vs. Non-AI help and AI help vs. Control (*Study* 2) Note: N = 1,203. Error bars represent 95% CI. Correlations among variables range from | *r* | = 0.05 to 0.76. Positive *d* values indicate higher values in the *AI Help* condition while negative *d* values indicate lower values in the *AI Help* condition.

age, and type of help received at work (*Help Conditions:* AI help, Non-AI help, or Control). The support described in the *Help Condition* varied based on the occupation. For example, in the *AI help* condition, a lawyer "sometimes asks generative AI to summarize information for (him/her)." In the *Non-AI help* condition, the lawyer "sometimes asks a paralegal to summarize information for (him/her)." For each occupation, the help provided in the *AI help* and *Non-AI* help conditions was described identically and only the source varied. In the *Control* condition, no statement about help was included. Given the effort-reducing potential of AI, we predicted that observers would be likely to evaluate people who use AI as lazy relative to people described as receiving other forms of help or people who are not described as receiving help.

We conducted *t tests* to evaluate the differences between assessments of laziness as well as the six measures of agency in the *AI Help, Non-AI Help,* and *Control* conditions. Fig. 2 displays a summary of the effect sizes for each dependent variable, with the effect size (*Cohen's d)* for the comparison between *AI Help* and *Non-AI Help* represented by a blue circle and the comparison between *AI Help* and *No Help* represented by a red triangle. The results revealed negative evaluations of employees who receive help from AI sources relative to those who receive help from non-AI sources or no help at all.

Consistent with our preregistered hypothesis, targets described as receiving help from AI were rated as lazier (M = 2.50, SD = 1.32) than employees described as receiving help from non-AI sources (M = 2.16, SD = 1.19, $t(787.5) = 3.79$, $P < 0.001$) or who were not described as receiving any help (M = 2.02, SD = 1.10, $t(766.9) = 5.55$, $P < 0.001$).

We also performed *t* tests on the six measures of agency and found that targets who received help from AI were also rated as less competent, less diligent, less independent, and less self-assured than targets who received help from other sources or received no help at all. The 95% CI indicate that the differences for laziness as well as four components of agency (competence, diligence, independence, and self-assuredness) were all significantly different from 0. Differences for the two remaining dimensions of agency—ambitious and dominant—were not significant. These results suggest that penalties for AI use arise in perceptions of effort and ability, but not those related to interpersonal competitiveness.

Testing a broad range of stimuli enabled us to examine whether the target's age, gender, or occupation qualifies the effect of receiving help from AI on these evaluations. We found that none of these target demographic attributes influences the effect of receiving AI help on perceptions of laziness, diligence, competence, independence, or self-assuredness.* This suggests that the social stigmatization of AI use is not limited to its use among particular demographic groups. The result appears to be a general one.

Study 3. Next, we examined whether people act on their beliefs about people who use AI in a two-stage incentive-compatible hiring task. In the first stage, we recruited 801 online participants (job *candidates*) who completed a short online task in which they had to count different colored tiles in a grid. After the task, they reported how often they use generative AI tools. In the second stage, we recruited 1,718 online participants to serve as *managers.* Managers were instructed that they were hiring for a task that involved describing the content of images and that they would be compensated based on how well the candidate they hired actually performed the task. Managers then reviewed the profile of a single candidate who had actually completed the task, which corresponded with a candidate who either uses AI regularly (Daily) or never (None).† Managers then responded to three items evaluating the employee's *Task Fit* and made a binary choice to either *Hire* the focal employee or instead hire a candidate randomly drawn from the pool of candidates who had completed the task. Afterward, they reported their own AI use.

We first tested the effects of manager AI use, candidate AI use, and their interaction on managers' evaluation of the candidate's *Task Fit* using a linear regression model (*SI Appendix*, Table S5)‡. On average, manager participants rated candidates who regularly use AI and candidates who never use AI as similarly fit for the task (P = 0.516). However, the effect of candidate AI use on *Task Fit* was dependent on the manager participant's own AI use (b = 0.06, SE = 0.01, $t(1664) = 7.08$, $P < 0.001$). Compared to those who use AI less frequently, manager participants who use AI more frequently tended to rate the candidate who uses AI daily as more fit for the task.

The results for the hiring decision were consistent with this pattern. The tendency to hire candidates in each condition depended on the manager participant's own AI use, as revealed by the interaction term in the logistic regression model shown in *SI Appendix*, Table S2 (b = 0.15, SE = 0.03, $z = 5.80$, $P < 0.001$). The model reveals that manager participants who use AI less frequently themselves favored the candidate that does not use AI at all, whereas manager participants who use AI more frequently favored the candidate that uses AI daily. Fig. 3 depicts these interactive effects.

One potential explanation for this finding is that people who use AI frequently are aware of the productivity gains it can afford and therefore expect that the candidates who use AI had an advantage completing the task. Because the description of the task did not explicitly state whether AI could be used, it is possible that high AI users inferred that AI could be used and the candidate who uses AI would perform well at the task. A second potential explanation is that dispositional inferences about AI users vary based on one's own AI use, and thus non-AI users may be more likely to penalize AI users in hiring decisions. We aim to shed light on both of these potential explanations in *Study* 4.

Study 4. In our final study, we examined whether perceptions of laziness mediate the relationship between candidate AI use and candidate outcomes. We also examined whether the social evaluation penalty could be offset if AI is clearly described as useful for the task for which the candidate is being considered.

*The *t* test results and models examining interactions for all measures of agency are provided in the *SI Appendix*.

†There were no differences in task performance based on candidate AI use (*SI Appendix*, Table S4).

‡In the model, Candidate Uses AI Daily was coded as 1, Candidate Never Uses AI was coded as -1, and manager participants' AI use was mean-centered.

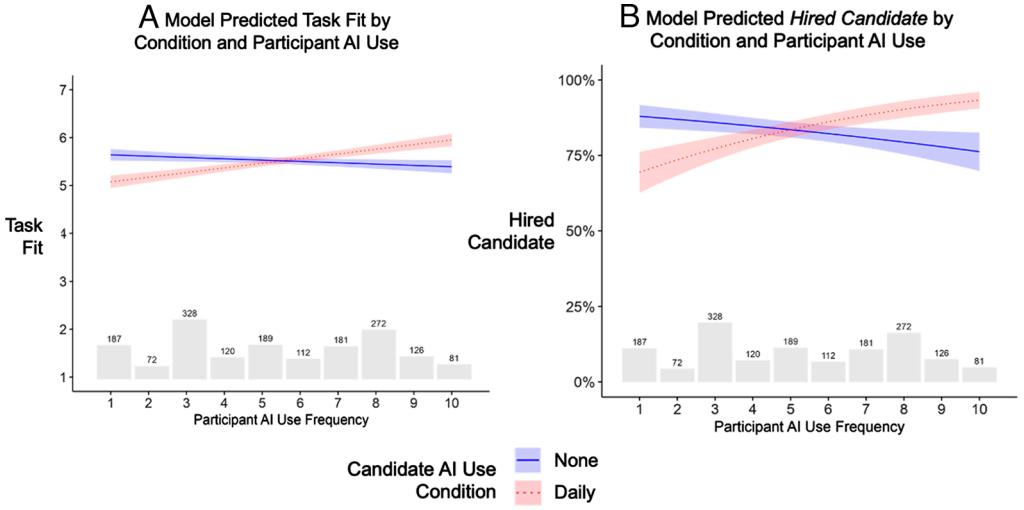

Fig. 3. The interactive effect of Candidate AI use (None vs. Daily) and Participant AI use on task fit and decision to hire candidate (*Study* 3) Note: N = 1,667. Panel A plots the model-predicted ratings for *Task Fit* (1–7 scale). Panel B plots the model-predicted probability that participants *Hired the Candidate* (1 = yes, 0 = no). Shaded regions represent 95% CI. Participant AI use frequency categories range from 1 (Never) to 10 (Multiple times per day). The histogram displays the number of participants in each Participant AI Use Frequency category.

To test these predictions, we recruited 1,006 online participants to complete a study in which they imagined they were hiring a gig worker for a task. They then read about a worker who either uses generative AI tools regularly at work (*AI Tool* condition) or a worker who regularly uses Microsoft Office programs at work (*Non-AI Tool* condition) and rated how lazy they perceive this worker to be. Next, they read that the task for which they were hiring was either a manual task (*Manual Task* condition: writing handwritten notes) or a digital task for which AI could be used (*Digital Task* condition: writing personal email messages). After completing the study, participants reported the frequency of their own AI use on a 1–5 scale. We predicted that candidates who use AI would be evaluated as lazier and that this social penalty would explain differences in perceptions of *Task Fit* and lower rates of *Hiring*. We also predicted that the relationship between AI use and these outcomes would be moderated by AI's utility for the task. Specifically, AI users would no longer be perceived as a poor fit for a task compared to non-AI users when using AI is possible and beneficial on the task.

We examined the effects of the *Tool* condition, *Task* condition, and their interaction on evaluations of *Task Fit.* A two-way ANOVA revealed significant main effects of both *Tool* condition ($F(1, 998) = 7.43$, $P = 0.007$) and *Task* condition [$F(1, 998) = 256.28$, $P < 0.001$], as well as a significant interaction between these variables [$F(1, 998) = 24.68$, $P < 0.001$]. For manual tasks, candidates using traditional tools were rated as having significantly higher task fit (M = 4.75, SD = 1.14) than those using AI tools (M = 4.18, SD = 1.50; $t(998) = 5.44$, $P < 0.001$). However, for digital tasks, there was no significant difference in task fit ratings between candidates using traditional tools (M = 5.58, SD = 0.98) and those using AI tools (M = 5.75, SD = 1.05; $t(998) = 1.59$, $P = 0.112$).

To identify whether *Laziness Perceptions* mediated the effect of *Tool* condition on *Task Fit,* we conducted a mediation analysis using the PROCESS Macro (Model 15, 5,000 bootstrapped resamples, (28)).[§] This analysis revealed that candidates using AI tools were perceived as lazier than those using traditional tools ($a = 0.47$, SE = 0.07, $P < 0.001$), and increased perceptions of laziness were associated with lower *Task Fit* evaluations. However, the effect of *Laziness Perceptions* on *Task Fit* was moderated by *Task*, with a stronger negative effect for digital tasks ($b = -0.43$, SE = 0.05, $P < 0.001$) compared to manual tasks ($b = -0.21$, SE = 0.05, $P < 0.001$). Nonetheless, the indirect effect of *Tool* on task fit through *Laziness Perceptions* was significant for both *Task* types.[¶] After accounting for this mediating effect of perceived laziness, the direct effect of tool choice on task fit depended on task type. For the manual task, using AI tools had a negative direct effect on *Task Fit* ($c' = -0.49$, SE = 0.10, $P < 0.001$). In contrast, for the digital task, using AI tools had a significant *positive* direct effect on *Task Fit* ($c' = 0.39$, SE = 0.10, $P < 0.001$). Taken together, these results show that AI users are perceived as lazier, which in turn is associated with lower judgments of task fit. However, this penalty was entirely offset for the Digital task for which AI was described as useful.

In light of our findings in *Study* 3 that participants' personal AI use shapes how they perceive AI users, we examined a separate mediation model testing whether the association between *Tool* condition and *Laziness Perceptions* was weaker for participants who reported higher levels of AI use [PROCESS Model 28, 5,000 bootstrapped resamples, (28)]. We found that the effect of candidate AI use on perceptions of laziness was moderated by the participants' personal AI use, such that participants who used AI more frequently were less likely to perceive the candidate who uses AI as lazy (*Personal AI Use* × *Tool*: b = −0.23, $P < 0.001$). To further probe this relationship, we separately analyzed *Laziness Perceptions* by *Tool* within each of the five levels of self-reported AI frequency (1 = has never used AI, 2 = tried AI once or twice, 3 = occasional use of AI, 4 = weekly use of AI, 5 = daily use of AI). At low to moderate levels of personal AI use (1–3 on the scale), the candidate who used AI tools was perceived as significantly lazier than the candidate who used non-AI tools (*SI Appendix*, Fig. S2). However, this difference was not significant for participants with high levels of AI use (weekly or daily). Fig. 4 depicts this moderated mediation model for *Manual Tasks* (Panel A) and *Digital Tasks* (Panel B), with one *a* path coefficient corresponding to *Low/Moderate Personal AI Use* (less than weekly) and a second *a* path coefficient corresponding to *High Personal AI Use* (weekly or more frequently).

We examined the same set of mediation models for the binary *Hiring* dependent variable. Directionally, all relationships were

[§]In our preregistration, we reported that the primary test of our hypotheses would be a mediation model with only the *c* path moderated by task (PROCESS, Model 5) but that, for exploratory purposes, we would separately examine task as a moderator of the *b* path (PROCESS, Model 15). We report the preregistered analysis in the *SI Appendix*, Fig. S3.

[¶]The index of moderated mediation was significant (Index: −0.10, SE = 0.04, 95% CI [−0.19, −0.03]), indicating that the indirect effect was stronger for digital tasks compared to manual tasks.

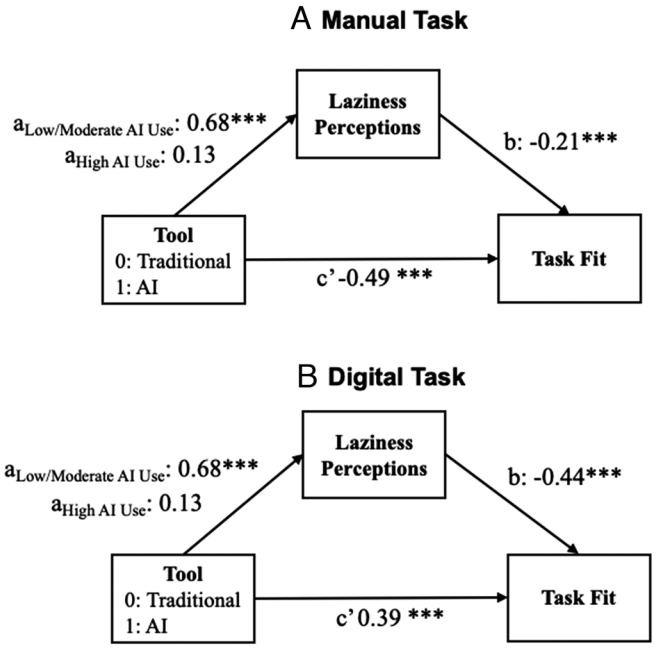

Fig. 4. Perceptions of laziness as a mediator between candidate tool use and task fit (*Study* 4) Note: N = 1,002. Coefficients are unstandardized. Panel (*A*) depicts the mediating effect of Laziness perceptions on the relationship between Tool and Task Fit for the manual task. Panel (*B*) depicts the mediating effect of Laziness perceptions on the relationship between Tool and Task Fit for the digital task. High AI Use refers to participants who use AI at least weekly. Low/Moderate AI Use refers to participants who do not use AI or who use AI less than weekly.

consistent with those we found for *Task Fit*; however, the interaction between *Tool* and *Participant AI use* did not quite reach statistical significance (P = 0.103). These models are reported in *SI Appendix*.

Overall, our findings suggest that the consequences of AI use for assessments of task fit are contingent on the evaluators' own use of AI tools and the task itself. Evaluators who do not use AI regularly (less than weekly) evaluate candidates who do use AI as lazy, and this translates to lower perceptions of task fit. However, this social penalty can be offset for tasks for which AI is clearly useful. When AI is described as being useful for the task, candidates who use AI are perceived as having higher *Task Fit.* However, when AI is not useful for the task, candidates who use AI experience a *Task Fit* penalty relative to those who use non-AI tools.

Discussion

The experiments reported in this research support our core prediction that individuals expect and receive a social evaluation penalty for receiving assistance from AI. *Study* 1 demonstrated that people believe others will perceive them negatively if they are known to receive help from AI, and *Study* 2 provided evidence that their hesitation to reveal AI use is well founded: evaluators tend to assess people who get help from AI as lazier, less competent, and less diligent than people who get help from other sources or who get no help at all. *Study* 3 showed that people who are not AI users themselves may act on their biases against employees who use AI by hiring them at a lower rate than employees who do not use AI. Finally, *Study* 4 revealed that the relationship between AI use and job assessment outcomes is mediated by perceptions of laziness. It also provided evidence of two contingencies. The association between AI use and laziness perceptions is dependent upon the evaluators' own AI use, and perceptions of laziness can be offset for tasks in which AI is clearly useful.

Implications. This work demonstrated that the use of productivity-enhancing tools can, paradoxically, erode social evaluations of their operators' competence and motivation. The implications of this work are relevant to at least three ongoing conversations in the psychology and organizational literatures. First, this work extends attribution theory by demonstrating how the source of assistance—not just its presence—shapes dispositional inferences about people who receive help. While prior work shows that receiving help can raise question about ability and motivation (8, 29), we demonstrate that agentic technologies like AI evoke strong dispositional attributions compared to more conventional forms of assistance. Our research examines people's beliefs about how they expect to be perceived for using help from different forms of *technology* and how others' evaluations are influenced by the type of help received. Our findings suggest that attribution processes are sensitive not only to whether help is received but also to observers' familiarity with and beliefs about the help source. This insight helps explain why people who achieve productivity gains through AI might paradoxically be perceived as less competent or motivated.

Second, this work offers insight into challenges surrounding AI adoption. Given the potential for AI use to enhance outcomes such as task performance (9), creativity (10), and decision-making (30), scholars have devoted considerable effort to understanding why people are reluctant to use it. This line of inquiry has largely focused on how potential AI users perceive the technology itself (e.g., 31–33), however, it has not explored the *social evaluation* consequences of AI use. This is an important oversight because people care about how their actions will be perceived by others (7): people may choose not to use AI—or not to disclose their use of AI—if they expect to incur a social penalty. We propose that this social evaluation penalty is an overlooked barrier to AI adoption.

Finally, this research enriches our understanding of how people evaluate unfamiliar technologies in organizational settings (34). Prior work suggests that people are especially likely to underestimate the impact of technologies that threaten the status quo (35). Our findings extend this literature by demonstrating that such skepticism extends beyond the technologies themselves to judgments about their users.

Limitations and Future Directions. Several limitations of this research warrant discussion. First, although our studies examined both anticipated and actual evaluations of AI tool use across different workplace contexts, all of our data were collected through experimental paradigms using online convenience samples. This approach afforded us tight control over variables of interest, but future research is necessary to demonstrate whether our findings generalize to social evaluations, task assignments, and performance assessments in established organizations.

Second, the operationalization of AI tools throughout our studies was intentionally general to capture broad perceptions. However, this approach may not have fully captured the nuanced ways in which people evaluate use of different types of AI systems in real-world contexts. For example, perceptions of people who use AI may depend on whether the tool augments existing work processes or automates entire tasks. Similarly, people may be less inclined to evaluate those who use AI negatively when the AI system is embedded within a more familiar technology (for example, an AI-enabled editing tool included in a word processor) (36). Although our stimuli in *Study* 2 examined AI use across a wide

range of organizational contexts and tasks, future research could systematically examine how the type of support AI provides shapes social evaluations of its users.

Finally, our studies were conducted during a period of rapid evolution in workplace AI capabilities and adoption (March 2024 to February 2025) (37). Perceptions of technology tend to be influenced by its age (35), and thus the social evaluation penalties we documented are likely to shift as AI tools become more commonplace and organizational norms around their use continue to develop. In a supplementary study (reported in *SI Appendix*, Fig. S1), we found no differences in expected or actual social penalties when AI use was described as rare (i.e., only one person uses it) or common (i.e., many colleagues also use it). Yet, the interaction effects we observed between evaluators' own AI use and their willingness to hire others who use AI in *Study* 3 and perceptions of others' laziness in *Study* 4 may foreshadow that the stigma we document may naturally decrease as more people gain direct experience with AI tools.

Conclusion

This work provides experimental evidence that people incur a social evaluation penalty for using AI tools at work. This generates a dilemma for employees: The productivity gains they can achieve with AI tools carry a social cost.

Materials and Methods

For each study, all data, analysis code, and study materials and preregistrations are accessible on the Open Science Framework (OSF) (38) https://tinyurl.com/yvrvsfvw. All studies were preregistered and approved by the Institutional Review Board at Duke University. All participants provided informed consent.

Study 1. We recruited 500 participants on the participant platform Prolific to complete this study.[#] We excluded three participants from our analysis based on our preregistered exclusion criteria, which stated that we would only analyze observations from participants who spent at least 30 s on the study and who did not provide identical responses to every rating item. Our final sample thus included 497 participants (M_{Age} = 37.7, 50% male). Nearly all participants (97%) held at least a bachelor's degree or higher. Participants were paid $0.60 for their time.

Participants were asked to imagine that they worked in the operations department of a company. Participants read that part of their duties as operations specialists included creating reports for customers, which was a manual process that required customizing each report and the information presented for each customer. Next, participants read that they had recently found a tool that could help them generate the reports much faster. Participants were randomly assigned to read that the tool was a generative AI tool (*AI Tool* condition) or a "dashboard creation tool" (*Non-AI Tool* condition). In both conditions, participants read that there were no rules prohibiting the use of this tool in their workplace and that they still ensure that the reports are accurate and meet quality standards.

After reading about the scenario, participants provided ratings on three items indicating whether they would disclose the use of this tool to their managers ($\alpha = 0.91$) and an additional three items indicating whether they would disclose the use of this tool to their colleagues ($\alpha = 0.93$). Next, participants were asked to rate how they believed others would view them on two dimensions of agency (27), competence ($\alpha = 0.96$), and diligence ($\alpha = 0.93$), as well as whether they believed others would perceive them as lazy ($\alpha = 0.92$) and replaceable ($\alpha = 0.95$).

Study 2. We recruited 1,215 participants from the Cloud Research Connect participant platform to complete this study. We excluded 12 participants based on our preregistered exclusion criteria, which stated participants who provided identical ratings to every scale item would be excluded. Our final sample was thus 1,203 (M_{age} = 39.1, 59% male). Most participants (69%) held at least a bachelor's degree. Participants were paid $0.75 for their time.

Participants read a paragraph describing an employee and were informed that they would be asked to answer several questions about their impressions of the employee. The paragraph presented was randomly selected from 384 unique stimuli that systematically manipulated several attributes: gender (male or female), occupation (lawyer, accountant, financial analyst, human resources, sales, software engineer, teacher, or consultant), age (25, 30, 35, 40, 45, 50, 55, or 60), and type of help received at work (*Help Conditions:* AI help, Non-AI help, or Control). The specific support described in the *Help Condition* varied based on the occupation. The full text of each of the stimuli is provided on our OSF site.[‖]

On the next screen, participants were asked to rate the extent to which they agree or disagree that several adjectives accurately describe the target employee. The employee's description was repeated on this page for reference. Participants completed scales rating the target's laziness ($\alpha = 0.95$), as well as six dimensions of agency (27): competent ($\alpha = 0.90$), diligent ($\alpha = 0.90$), ambitious ($\alpha = 0.81$), independent ($\alpha = 0.92$), self-assured ($\alpha = 0.89$), and dominant ($\alpha = 0.93$).

Study 3. We collected data for this study from two separate participant pools: the *candidate* pool and the *manager* pool. First, we recruited 801 *candidate* participants from the Cloud Research Connect platform to complete a task. These participants comprise the candidate pool, some of whom would ultimately be presented to the manager participants as stimuli in the candidate hiring task. We attempted to constrain the variation in the employee participants' demographics by recruiting only participants aged between 30 and 45 who held at least a bachelor's degree. The purpose for these restrictions was so that we could embed the information about AI use presented to the manager participants alongside other characteristics that were accurate but consistent across candidates. The candidate participants were paid $0.40 for completing the task and a bonus of up to $0.10 for performance on the task.

Second, we recruited 1,718 *manager* participants (M_{age} = 38.5, 54% female). Most participants (62.4%) held at least a bachelor's degree. We excluded 50 participants based on our preregistered exclusion criteria, which required that participants pass a comprehension check and spend at least 45 s completing the study. Our final sample of manager participants was thus 1,668.

Candidate participants began by reading a set of instructions for a task referred to as the "tile counting" task. For this task, candidate participants were asked to review 10 × 10 grids that contained a mixture of green, red, and gray tiles. They were then asked to rank the colors based on the frequency with which they appeared in the grid. The task was designed to rely primarily on effort rather than ability. Candidate participants were informed they could review up to 10 images total and that they would have 90 s to review as many of the grids as they could. We also informed them that they could receive a bonus of up to $0.10 for accuracy. When candidate participants advanced to the page following the instructions, a 90 s countdown timer started automatically and was displayed at the top of the page. When the timer finished, the screen automatically advanced to the next page, which contained demographic questions and a question about AI use. Specifically, we asked them how often they use AI tools (such as ChatGPT, Claude, and Gemini).

Several days later, we recruited manager participants to complete the study. Manager participants were informed that they would be in the role of manager for this study and that their task would be to hire a candidate to perform a particular task. They then read that the task for which they were hiring requires the candidate to carefully examine 10 images and make judgments about their content and that the accuracy of those judgments would then be evaluated based on objective criteria. To ensure incentive compatibility, we informed manager participants that all candidates had already completed the task and that they as managers would be eligible to earn a bonus of up to $0.25 based on how the candidate they selected actually performed on the task.

Manager participants read that they would be presented with a single candidate from the pool of potential employees and that they would have the option of hiring *this* candidate to complete the task or hiring a randomly selected candidate from the remaining candidate pool. The purpose of this binary choice was so that participants could be compensated based on the actual performance of one candidate. Manager participants reviewed five data points for the candidate presented, three of which were held constant across all stimuli: age range (30 to 45), highest level of education (bachelor's degree or higher), and employment status (full-time, part-time, or business owner). We varied two characteristics of the candidate: gender (male or female) and

[#]We replicated the results of *Study* 1 on Pollfish, a market research platform. The effect sizes for this replication study were similar and are reported in the *SI Appendix*, Table S1.

[‖]There was a typo (an incorrect pronoun) in 40 of the 384 stimuli. This error occurred in both the AI Help and Non-AI Help conditions, and we confirmed that our conclusions are identical if we exclude all of the data collected with each of these stimuli.

generative AI use (e.g., ChatGPT) (Daily or None). We used a random number generator within the survey platform (Qualtrics) to match each manager participant with a real candidate that corresponded with the candidate profile that was displayed to them.

After reviewing the candidate's information, manager participants evaluated the candidate's *Task Fit* using a three-item scale ($\alpha = 0.90$) and indicated their choice to hire the candidate they reviewed for the task or to hire a randomly selected candidate from the remaining candidate pool for the task. After making their selection decision, participants reported their own AI use frequency (on a 1 to 10 scale) and supplied demographic information. After the study, we awarded bonuses of up to $0.25 based on how the worker the manager participant hired performed on the task. Manager participants who opted to hire the candidate whose information they reviewed were paid based on the performance of the candidate they had been matched to, and we matched manager participants who chose to hire a random candidate to a random candidate from the remaining pool.

Study 4. We recruited 1,006 participants from the Cloud Research Connect participant platform to complete this study ($M_{age} = 43.2$, 57% female). Nearly all participants in the sample (98.6%) held at least a bachelor's degree. We excluded 3 participants based on our preregistered exclusion criteria, which stated we would exclude participants that provided identical ratings to every scale item. We also excluded 1 additional participant who skipped the three items measuring task fit. Our final sample was thus 1,002. Participants were compensated $0.60 for their time.

Participants were asked to imagine working at a nonprofit organization that was hiring a "gig worker" for a short-term task. They read a paragraph describing a candidate named Mike, a customer support representative applying for the position. Participants were randomly assigned to either the *AI Tool* condition, where Mike was described as frequently using AI tools to organize information, prepare reports, and develop customer training guides, or the *Traditional Tool* condition, where he used Microsoft Word and PowerPoint for these same tasks. In both conditions, Mike was described as available to start as soon as that weekend. After reading about the candidate, participants rated how lazy they perceived him to be using four items ($\alpha = 0.94$). We measured perceived laziness before introducing the specific task details to capture participants' general impressions of the candidate independent of the task requirements.

On the next page, participants read more detail about task. The text explained that the nonprofit was hiring a temporary worker to help write personal thank you messages to 750 donors. Participants were randomly assigned to read that this was a *Manual Task* or a *Digital Task*. In the *Manual Task* condition, the task was described as writing 750 handwritten notes and mailing them to the donors' home addresses. In the *Digital Task* condition, the task was described as writing 750 personal email messages. To ensure it was clear that AI tools would be useful for the digital task and available to the worker, we added the following statement to the end of the description: "Generative AI tools might be useful for this task, and the worker is welcome to use them." The worker candidate's description was then repeated on this page for reference. Participants then responded to three items measuring the candidate's *Task Fit* ($\alpha = 0.92$) and make a yes-or-no recommendation as to whether the organization should *Hire* the candidate for the task.

Data, Materials, and Software Availability. Data for these experimental studies have been deposited in OSF (https://osf.io/beun6/) (38).

ACKNOWLEDGMENTS. We gratefully acknowledge research funding from the Fuqua School of Business at Duke University.

1. F. Heider, *The Psychology of Interpersonal Relations* (Wiley, 1958).
2. E. E. Jones, K. E. Davis, From acts to dispositions: The attribution process in person perception. *Adv. Exp. Soc. Psychol.* **2**, 219–266 (1965).
3. B. Weiner, *An Attributional Theory of Motivation and Emotion* (Springer-Verlag, 1986).
4. B. F. Malle, "Attribution theories: How people make sense of behavior" in *Theories in Social Psychology*, 2nd Ed., pp. 93–120. (2022).
5. L. Ross, R. E. Nisbett, *The Person and The Situation: Perspectives of Social Psychology* (Pinter & Martin Publishers, 2011).
6. L. Ross, The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Adv. Exp. Soc. Psychol.* **10**, 173–220 (1977).
7. E. Goffman, *The Presentation of Self in Everyday Life* (Doubleday, 1959).
8. D. T. Gilbert, D. H. Silvera, Overhelping. *J. Pers. Soc. Psychol.* **70**, 678 (1996).
9. S. Noy, W. Zhang, Experimental evidence on the productivity effects of generative artificial intelligence. *Science* **381**, 187–192 (2023).
10. N. Jia, X. Luo, Z. Fang, C. Liao, When and how artificial intelligence augments employee creativity. *Acad. Manag. J.* **67**, 5–32 (2024).
11. Asana, *State of AI at Work. Asana Work Innovation Lab*. https://asana.com/resources/state-of-ai-work (2024).
12. R. Heath, Exclusive: Employees are bringing their own AI to work. *Axios*. (2024), https://www.axios.com/2024/05/08/employees-bring-their-own-ai-microsoft-linkedin.
13. B. Marr, The employees secretly using AI at work. *Forbes*. (2024), https://www.forbes.com/sites/bernardmarr/2024/09/05/the-employees-secretly-leveraging-ai-at-work/.
14. D. S. Werner, *Myth and philosophy in Plato's Phaedrus* (Cambridge University Press, 2012).
15. R. Hembree, D. J. Dessart, Effects of hand-held calculators in precollege mathematics education: A meta-analysis. *J. Res. Math. Educ.* **17**, 83–99 (1986).
16. H. R. Arkes, V. A. Shaffer, M. A. Medow, Patients derogate physicians who use a computer-assisted diagnostic aid. *Med. Decis. Making* **27**, 189–202 (2007).
17. B. S. Vanneste, P. Puranam, Artificial intelligence, trust, and perceptions of agency. *Acad. Manag. Rev.* **42**, 1–46 (2024).
18. A. S. DeNisi, K. R. Murphy, Performance appraisal and performance management: 100 years of progress? *J. Appl. Psychol.* **102**, 421 (2017).
19. M. Inzlicht, A. Shenhav, C. Y. Olivola, The effort paradox: Effort is both costly and valued. *Trends Cogn. Sci.* **22**, 337–349 (2018).
20. M. Bolino, D. Long, W. Turnley, Impression management in organizations: Critical questions, answers, and areas for future research. *Annu. Rev. Organ. Psychol. Organ. Behav.* **3**, 377–406 (2016).
21. M. B. Kay, D. Proudfoot, R. P. Larrick, There's no team in I: How observers perceive individual creativity in a team setting. *J. Appl. Psychol.* **103**, 432 (2018).
22. D. A. Lagnado, T. Gerstenberg, R. I. Zultan, Causal responsibility and counterfactuals. *Cogn. Sci.* **37**, 1036–1073 (2014).
23. Y. Xiang, J. Landy, F. A. Cushman, N. Vélez, S. J. Gershman, Actual and counterfactual effort contribute to responsibility attributions in collaborative tasks. *Cognition* **241**, 105609 (2023).
24. H. H. Kelley, The processes of causal attribution. *Am. Psychol.* **28**, 107 (1973).
25. A. S. Rosette, J. S. Mueller, R. D. Lebel, Are male leaders penalized for seeking help? The influence of gender and asking behaviors on competence perceptions *Leadership Q.* **26**, 749–762 (2015).
26. E. Glikson, A. W. Woolley, Human trust in artificial intelligence: Review of empirical research. *Acad. Manag. Ann.* **14**, 627–660 (2020).
27. A. Ma, A. S. Rosette, C. Z. Koval, Reconciling female agentic advantage and disadvantage with the CADDIS measure of agency. *J. Appl. Psychol.* **107**, 2115 (2022).
28. A. F. Hayes, Beyond baron and kenny: Statistical mediation analysis in the new millennium. *Commun. Monogr.* **75**, 408–420 (2008).
29. P. S. Thompson, M. C. Bolino, Negative beliefs about accepting coworker help: Implications for employee attitudes, job performance, and reputation. *J. Appl. Psychol.* **103**, 842 (2018).
30. M. H. Jarrahi, Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Bus. Horiz.* **61**, 577–586 (2018).
31. B. J. Dietvorst, J. P. Simmons, C. Massey, Algorithm aversion: People erroneously avoid algorithms after seeing them err. *J. Exp. Psychol. Gen.* **144**, 114 (2015).
32. M. Li, T. B. Bitterly, How perceived lack of benevolence harms trust of artificial intelligence management. *J. Appl. Psychol.* **109**, 1794–1816 (2024).
33. K. C. Yam *et al.*, Robots at work: People prefer–and forgive–service robots with perceived feelings. *J. Appl. Psychol.* **106**, 1557 (2021).
34. A. Orben, The Sisyphean cycle of technology panics. *Perspect. Psychol. Sci.* **15**, 1143–1157 (2020).
35. A. H. Smiley, M. Fisher, The golden age is behind us: How the status quo impacts the evaluation of technology. *Psychol. Sci.* **33**, 1605–1614 (2022).
36. R. S. Rao, R. K. Chandy, J. C. Prabhu, The fruits of legitimacy: Why some new ventures gain more from innovation than others. *J. Marketing* **72**, 58–75 (2008).
37. M. Heikkila, W. D. Heavan, *What's Next for AI in 2024. MIT. Technology Review* (2024), https://www.technologyreview.com/2024/01/04/1086046/whats-next-for-ai-in-2024/.
38. J. A. Reif, R. P. Larrick, J. B. Soll, Data from "Experimental evidence of a social evaluation penalty for using AI". Open Science *Framework*. https://osf.io/beun6/. Deposited 20 March 2025.