Employee Preferences for Artificial Intelligence-driven Performance Evaluation Systems in a Post-COVID-19 Workspace

Jasmijn Bol[*]
Tulane University

Conor V. C. Brown
Grand Valley State University

Lisa LaViers[**]
Tulane University

May 2023

*The authors would like to thank the Institute of Management Accountants for funding this research. We would also like to thank Maximilian Margolin, Jordan Samet, Steven Salterio, Daniel Schaupp, Suyun Wu, Michael Yip, participants of the 2021 AAA Management Accounting Section conference, participants of the 2022 Labor and Accounting, ENEAR, and EXRIMA conferences and workshop participants at Tulane and Monash Universities for their helpful comments on this paper.

[**]Corresponding Author: Please contact at llaviers@tulane.edu

ABSTRACT:

The COVID-19 pandemic ushered in a new age for the workplace with more environmental instability and remote work, which led many firms to rethink their performance evaluation systems. At the same time, technological improvements have allowed AI-driven performance evaluation systems to become a realistic alternative to human-driven systems for an increasing number of jobs. As a result of these simultaneous changes, firms are increasingly interested in making the transition to AI-driven systems. However, prior research suggests that employees are concerned about the use of AI, because they believe that AI does not consider the context in which performance occurs. We argue that employee preferences, which are critical to the motivational impact of the incentive system, are not uniform. Specifically, we predict and find that employee preferences for AI are relatively weaker when the business environment is unstable. We predict and find that employees who have experienced workplace discrimination in the past have a higher relative preference for an AI-driven system, and this preference is stronger when working remotely than when working in a shared workspace. When working in a shared workspace, employees with lower social intelligence have a stronger relative preference for AI-driven systems than employees with higher social intelligence. We find that this effect disappears when employees work remotely. These findings help guide firms that are considering investing in AI-driven performance evaluation systems.

KEYWORDS: Performance Evaluations, Artificial Intelligence, Objective Performance Evaluation, Subjective Performance Evaluation, Discrimination, Social Intelligence, Remote work

I. INTRODUCTION

In March 2020, the global COVID-19 pandemic spurred a transformation of the modern workplace. As a result, firms were forced to make several changes to the way employee performance is evaluated (Knight 2020, Lund et. al 2021). First, the pandemic led to dramatically increased environmental instability that even years later has not resolved (Baker et al. 2020). This environmental instability makes performance evaluation more difficult for managers because much of the historical performance data from before and during the pandemic is no longer useful as a benchmark to assess employee performance. Second, more employees began working remotely, and consequently, the opportunities for supervisors to rely on informal observations and conversations as part of their employee performance assessment became more limited (Knight 2020; Mackenzie et al 2020). Instead of in-person observations, managers must make assessments based on digital interactions with employees. The amount of informal performance-irrelevant data was also significantly reduced.

These two major changes in the workplace happened simultaneously with a dramatic improvement in artificial intelligence (AI) (Kellogg, Valentine, and Christin 2020). Because of these improvements, AI-driven performance evaluation systems became a realistic alternative for traditional performance evaluation systems, which rely, at least partly, on human judgment (Baker, Gibbons, and Murphy 1994).[1] The improved technology in combination with the changes in the workplace spurred on by COVID-19 resulted in increased interest and investment in AI-driven systems (Knight 2020; Lund et. al 2021). It is, however, yet unclear how successful the systems will be. Part of why the success is difficult to predict is because it is not only dependent on how

[1] Judgement-based evaluation is also widely referred to as subjective evaluation (see e.g., Bol 2008).

technologically advanced the systems are, but also on how employees will react to having their performance evaluated by machines instead of by humans.

To contribute to the collective understanding of when transitioning to AI-driven performance evaluation may improve firm outcomes, we examine employees' relative preferences for AI-driven versus human-driven performance evaluation systems and the impact of two important pandemic-related factors on these preferences: environmental instability and remote work. Although the strong investment in AI clearly signals that top managers see potential in AI-driven systems, there is evidence that employees are less enthusiastic about the switch because they are concerned about how fair the AI-driven systems will be (Kellogg, Valentine, and Christin 2020).

These employee concerns are consistent with prior research on algorithm aversion, which finds that people are resistant to following the recommendations and judgments that AI makes *for* them as a decision aid (Dietvorst, Simmons, and Massey 2015; Burton, Stein, and Jenson 2020). Research examining how employees feel about AI making judgments *about* them is scarcer but also shows employee reluctance towards AI. Specifically, Newman, Fast, and Harmon (2020) argue that employees prefer human-driven systems to AI-driven systems because employees believe AI-driven systems decontextualize performance information more than humans. Decontextualization means the evaluator (i.e., a human manager or AI) considers only the performance metrics, not the context in which the performance occurred. In other words, employees believe AI makes an objective evaluation of them that is not sufficiently adjusted for factors outside of their control. Interestingly, Newman et al. (2020) find that these negative perceptions hold even when the two systems make the exact same performance assessment.

This finding about negative perceptions is important because for employees to put forth high levels of effort, they must believe that in the future the performance evaluation system will fairly compensate them for it (Masterson et al 2000; Bol 2011). Employees who believe that the evaluation systems of their organization are not fair have lower performance, lower effort, and are more likely to leave the organization compared to employees who believe that their organizations' systems are fair (Poon 2004, Kuvaas 2006). These results suggest that AI-driven systems may be less effective than anticipated simply because employees believe they are less fair than human-driven systems even if the systems make identical evaluations. Employee preferences and confidence in the fairness of the different systems are critical to effectiveness and should therefore be considered in implementation decisions.

By examining employee preferences for the different types of systems, we expand on Newman et. al's (2020) findings. We investigate whether employees' relative preferences for human-driven versus AI-driven performance evaluation systems are constant and how changes to the workplace spurred by the pandemic may have affected these relative preferences. Note that we are examining relative preferences when the human-driven and AI-driven systems have the same formal employee performance information. Any differences in preferences are therefore not driven by one system having access to performance-related data that the other does not but instead are driven by employee's beliefs about how the system will process that information.

First, we examine the impact of environmental stability on relative preferences. We argue that in unstable environments, employees will have a higher preference for judgment-based human-driven systems because employees will believe that a fair evaluation will require consideration of, and potential adjustments for, special circumstances and that human evaluators are better at making these adjustments. Consequently, we hypothesize that employees will have a relatively lower preference for AI-driven systems versus human-driven systems in unstable environments than in stable environments.

Second, we examine the impact of working remotely on employees' relative preferences. Although both systems have the same formal performance data on which to base their evaluations, human evaluators will also be influenced by informal performance-irrelevant information that they have gathered through observations and interactions. This information might include what the employee looks like or their favorite sports team. They will do this even though this type of information is not performance related and will have greater access to this type of information when working in a shared office with their employees than when working remotely away from the employees. Unlike human managers who will have differing levels of access to this information based on where the employees work, AI-driven systems will never have access to this type of information and cannot be swayed by it. Because there is a difference in how the systems are influenced by performance-irrelevant information, we posit that employees' relative preferences for AI-driven versus human-driven systems are influenced by whether employees work remotely or in a shared workspace.

However, we do not predict a uniform effect. Instead, we argue that when employees believe their performance evaluations will be negatively impacted by performance-irrelevant information, like employees who have experienced workplace discrimination in the past, they will have a relatively stronger preference for AI-driven systems. Consistent with intergroup contact theory of prejudice, we predict that this preference for AI-driven systems will be stronger when working remotely because limited informal social interactions make human biases like discrimination stronger (Pettigrew 1998). On the other hand, when employees believe they will benefit from the inclusion of performance-irrelevant factors in their assessments, like when they have high social intelligence and are gifted at workplace interactions such as "water cooler conversations," they will have a higher preference for human-driven systems. However, this preference will be reduced when employees work remotely due to the more limited opportunities to influence their managers' assessments through informal data.

We use two scenario-based experiments in an online labor market. Each experiment has a manipulation and measured variables. In both of our experiments, participants assume the role of employees at a hypothetical firm. After learning about the firm and nature of the performance evaluation, participants are asked to indicate their preference for a human-driven or AI-driven performance evaluation system. In experiment 1, we vary the environmental stability of the business. In experiment 2, we manipulate the employees' workspace; employees either work remotely and have lower opportunities for social contact or in a shared office with their managers and have higher opportunities for social contact. We also measure employees' perceptions of past workplace discrimination and social intelligence scores using the Tromsø Social Intelligence scale (Silvera, Martinussen, and Dahl 2001).

We find that employees' preferences for human-driven versus AI-driven systems are indeed not uniform. Employees show a significantly lower relative preference for AI-driven systems in an unstable environment than in a stable one. Through mediation analysis, we show that employees are more concerned about an AI-driven system's ability to fairly consider the impact of the environment in an unstable environment than in a stable environment. In contrast, their concerns about a human manager's ability to fairly consider the environment are unchanged between stable and unstable environments.

Using data from both experiments 1 and 2, we show that employees who feel they have faced workplace discrimination in the past have a relatively stronger preference for AI-driven systems than employees that have not faced workplace discrimination. We also find that the preference for AI-driven performance measurement systems among participants who believe they have experienced workplace discrimination is stronger when participants work remotely. Lastly, using data from experiment 2, we find that participants with higher social intelligence have a stronger relative preference for human-driven systems than participants with lower social intelligence and that this effect differs based on where employees work. When working remotely, social intelligence plays no significant role in preferences for AI-driven versus human-driven systems. When working in a shared office, however, those with higher social intelligence prefer human-driven systems. We extend these results in supplemental analysis. Using a moderated-mediation model, we show that the effect of social intelligence on preference for an AI-driven or human-driven evaluation system is driven by a difference in perceived fairness of the evaluator.

This study contributes to the management and accounting literature and to practice in several ways. First, our findings contribute to the theoretical knowledge base about subjective, judgment-based evaluation of performance versus objective measurement of performance. Because the advances in AI now allow for more complete objective performance measurement for many more jobs, many organizations are considering the switch to AI-driven systems. We highlight that employee preferences need to be considered in management's assessments of human-driven versus AI-driven systems because those preferences directly impact motivational effectiveness. Incorporating employee preferences, however, is not a straightforward task. We contribute to the literature by showing that employees' preferences for human-driven versus AI-driven performance evaluation systems are not uniform across all types of employees and in all environments. That is, we show that how employees perceive the AI-driven performance evaluation system is not just driven by a blanket effect of algorithm aversion, but also by employees' beliefs about whether a judgement-based or objective evaluation is fairer. In some cases, employees want a more subjective, judgment-based human-driven system that they perceive will take their economic environment into consideration, while in other cases, employees prefer an AI-driven system that they perceive only measures their performance and ignores other irrelevant factors.

Our study also contributes to the growing literature on differences in how to manage employees who work in a shared or remote workspace. Our findings suggest that the relative preference between the AI-driven versus human-driven systems depends on the opportunity for supervisors to informally observe employees. This insight is important because the percentage of professionals who have left the office to work remotely has increased significantly over the last decade, and due to the COVID-19 pandemic, this number increased even more rapidly in March 2020. In this new more remote world, firms might feel AI-driven systems are the natural next step because employee performance needs to be captured digitally regardless of which person or system will evaluate it. While some employees will likely welcome this approach, like those who have been discriminated against in the past, others will likely be warier, such as employees with high social intelligence. This highlights again that when considering a potential switch to an AI-driven system, management must also carefully consider the preferences of their workforce, because these preferences influence the motivational effect of the performance evaluation system and employee retention.

This research also contributes to the work of computer scientists who are studying AI development and algorithm aversion (Burton, Stein, and Jensen 2020). While these researchers are working to advance the technology, they lack a deep understanding of the corporate performance evaluation processes. By showing how AI functions in business environments, we better illustrate how end users will react to this technology and how it can be best implemented. We hope that by working together, we can gain a better understanding of where advancement of this technology might be most beneficial to organizations.

II. LITERATURE REVIEW

Research has shown that rewards linked to performance assessments can result in increased productivity (Engellandt and Riphahn 2011). However, prior studies have also shown that for performance evaluation systems to effectively motivate employees, employees need to believe that their effort will be fairly rewarded (Trevor, Reilly, and Gerhart 2012; Downes and Choi 2014). Employees are the first movers, that is, they need to accept the contract and provide the effort before their performance is evaluated and rewarded. As a result, if employees do not think that performance will be rewarded fairly, they will reduce their effort levels or leave the firm (Simons and Roberson 2003; Bol 2011). Organizations can implement elaborate systems to accurately capture performance, but if the employees do not accept the system, it will still not have the desired motivational effect (Trevor et al. 2012).

While a plethora of studies have been conducted on employee performance evaluation in general and the role of fairness perceptions specifically (Prendergast and Topel 1993; Jacob and Lefgren 2008; Demeré, Sedatole, and Woods 2019), almost all of these studies have focused on human-driven evaluation systems because, until recently, it was the only type of system possible for the large majority of organizations, because reasonably complete objective performance data was not available. As a result, the question of whether employees prefer judgment-based evaluation or objective measurement has not been the focus of the literature; the inclusion of supervisor judgment was not a choice but a necessity for most organizations. As AI-driven evaluation systems emerge as viable alternatives, the inclusion of supervisor judgment becomes a choice for many more organizations, and consequently, firms need to compare the advantages and disadvantages of the different systems. Considering the importance of employee acceptance of a system for it to effectively motivate performance, examining employee preferences and perceptions of AI-driven versus human-driven systems should be part of this overall analysis.

In this study, we start by discussing the advantages and disadvantages of each type of system and then develop hypotheses regarding employees' preferences. Note, though it is possible for AI and humans to work together on decision-making, in this research, we specifically examine the AI-driven performance evaluation systems where AI and humans act independently from one another, rather than a system in which a human manager uses an AI recommendation as a decision aid.

Human-Driven Performance Evaluation

Traditionally, managers have evaluated employee performance using a combination of objective measurements and their own judgement (Prendergast 1999; Bol 2008). For most jobs, the manager is asked to at least subjectively determine the weight placed on the different performance measures and/or adjust for uncontrollable or unanticipated factors when deemed appropriate (Prendergast and Topel 1993; Ittner, Larker and Meyer 2003). Often, however, the manager is also asked to make assessments on dimensions that cannot be measured easily, like leadership skills and the quality of project execution (Ittner, Larker and Meyer 2003). Allowing managers to have the discretion to apply their judgment can lead to more complete assessments of the employees' performance, skills, and long-term contributions (Fisher et al. 2005; Gibbs et al. 2005; Bol 2011).

While supervisor discretion can be advantageous because of its ability to make assessments more complete, it can also have a dark side. Managers can introduce bias into the evaluation process. In an international survey conducted by Glassdoor (2019), approximately 50% of respondents report having personally seen or experienced ageism, sexism, racism, homophobia or favoritism (e.g., preference based on supporting the same sports team or having similar tastes in music) in their workplace. These experiences occur because managers are subconsciously or consciously including informal non-performance related information about employees in the evaluation process. Because of these experiences, employees are concerned about the fairness of performance assessments that include judgement by a human manager (Chan and Dimauro 2020; Moise and Cruise 2020).

Beyond the problem of bias, the other major disadvantage to human-driven evaluation is costliness. The evaluation process is typically slow and labor intensive (Rogel 2020). Estimates of the average time spent on performance evaluations are as high as hundreds of hours per year (Cappelli and Tavis 2016). Before dropping annual performance reviews, Deloitte Inc. estimated that the company spent nearly two million hours each year on performance evaluations.

AI-driven Performance Evaluation

Technical progress in the field of AI has made it possible for companies to switch to AI-driven performance evaluation systems. Some firms are developing their own AI-driven systems for this purpose. For example, IBM uses its self-developed AI-driven system (called Watson) to predict future employee performance (Greene 2018). IBM claims this has resulted in large cost savings with a reduction of human resources staff by 30% (Rosenbaum 2019). Other firms like Enaible and ButterFly.Ai are selling turnkey AI-driven performance evaluation systems. While these AI-driven systems may make up only a small portion of all performance evaluations currently, industry experts predict that the trend will continue to grow (Holsinger et. al 2019).

AI-driven systems are attractive to managers because of the perceptions that AI could provide a simple, cost-effective, and efficient new way to conduct evaluations (Fecheyr-Lippens, Schaninger, and Tanner 2015; Cheng and Hackett 2021). AI allows firms to use objective performance measurement for more jobs than ever before. AI systems can now replace individual managers' judgements in two ways: it can make objective assessments of many different types of data (e.g., numerical analysis and textual analysis) and develop weights that consistently adjust across every employee for known performance influencing factors like economic conditions and competitors' actions. Because of its processing speed, it can do complex math faster than humans can and, unlike human managers, it does not suffer from cognitive overload that results in bias (Bol, Margolin, and Schaupp 2023). Further, AI can learn from new data and adapt an underlying set of complex algorithms to adjust to novel circumstances.

While management may be enthusiastic about AI-driven performance evaluation because of its potential to cut costs and streamline the evaluation process, employees' potential responses to it are less clear. There is some research on algorithm appreciation that suggests that people may prefer the recommendation of an algorithm to one from a human (Castelo, Bos, and Lehmann 2019; Berger et al. 2021). However, there is a larger body of research on algorithm aversion that finds that most people prefer to rely on guidance from a human advisor more than an algorithm (Dietvorst, Simmons, and Massey 2015; Prahl and Van Swol 2017; Burton, Stein, and Jenson 2020). Both literature streams mainly look at the acceptance of a recommendation by an algorithm. The research on how employees perceive evaluations by an AI is more limited. One of the small number of studies in this area, Newman et. al (2020), shows that employees perceive performance evaluations made by AI to be less fair than those made by human evaluators. They argue that this is caused in part by employees' belief that AI does not consider the context in which employees operate. Consistent with their predictions, the authors find that employees prefer human-driven systems, even when the outcomes are identical.

We contribute to this line of research by examining if employees' relative preference for AI-driven versus human-driven performance evaluation systems are constant. Moreover, we examine the impact of two important changes to the workplace spurred by COVID-19 on employees' relative preferences for performance evaluation systems: the stability of the economy and remote work.

As discussed above, in our research setting the human manager and AI are using the same set of formal performance measures to judge performance. Moreover, we assume that both systems are functional. The hypothesized differences in preferences are therefore predicted to come from employees' perceptions of how the information is processed and analyzed, and the inclusion of non-performance relevant informal information.

III. HYPOTHESIS DEVELOPMENT

Environmental Stability

When an organization is operating in a stable environment, the context in which employees' performance occurs is relatively unchanging. With many periods of comparable information, an organization can develop reasonable benchmarks and prediction models of not only employees' performance but also of other factors that are known to influence performance like economic factors and competitor actions. We predict that employees will perceive that the comparable historical data makes contextualization of their performance based only on objective data easier and that adjustments based on human-judgement are less important. Under these circumstances, employees are less concerned about potential decontextualization by objective AI-driven systems. Moreover, Hu (2021) finds that in stable environments, employees build trust in

AI-driven systems because of the consistency of their adjustments. In contrast, employees may doubt that managers will match this level of consistency in judgement in stable environments due to their cognitive limitations and tendency to bias (Prendergast and Topel 1993; Ittner, Larker, and Meyer 2003).

In an unstable environment, historical data may not be perceived by employees to be representative of the current context in which employees' performance occurs. Employees will likely feel that there are new factors that impact their performance which need to be contextualized for their assessment to be an accurate reflection of their effort. For example, we posit that during the COVID-19 pandemic, employees will prefer their performance to be assessed in the context of the new stay-at-home orders, not just the economic and competitive factors that were considered before the pandemic. As a result, AI's strict objectivity may make AI-driven performance evaluation less attractive to employees in unstable environments than in stable ones.

This prediction, however, is not without tension. It is not clear whether a human or an AI would *actually* perform better at performance evaluation under instability. While AI uses models explicitly built on historical data, human managers also have mental models of what "high performers" are like based on the same data (Bol and Leiby 2018). Thus, it is also not easy for human managers to fairly consider new contextual factors that influence performance. Moreover, humans may be cognitively overloaded by needing to make complex assessments using unfamiliar data sources or larger adjustments from their mental models (Simon 1990). When humans are cognitively overloaded, they are more likely to engage in biases and inaccuracies. AI will not be cognitively overloaded and may in fact be better at making the types of large adjustments needed to examine the data in the unstable period. Thus, it is not clear which system would better assess 'true' underlying performance.

Even though both human-driven and AI-driven evaluators will have a harder time evaluating performance in unstable environments, we predict that a lack of stability will increase the relative preference for human-driven versus AI-driven evaluation systems. We predict that this change in preferences is driven by employees' increased concerns about AI's objectivity. In contrast, when the environment is stable, we predict that employees will have a relatively higher preference for AI-driven systems because the historical data will allow the AI to perform objective evaluation in a consistent fashion. Thus, we hypothesize that the stability of the environment influences employees' preferences for human-driven versus AI-driven performance evaluation systems.

H1: Employees in unstable environments show relatively lower preferences for AI-driven systems versus human-driven systems than employees in stable environments.

Remote Work

Although we investigate preferences for human versus AI-driven systems when the human and the AI have the same performance data available, we know that human evaluations are also influenced, consciously or unconsciously, by information that is not performance relevant (Du, Tang, and Young 2012; Bol 2011). This type of data includes information like employees' physical appearance, whether an employee has a mutual hobby with their manger, or whether they are from the same region. This type of informal data is gathered by managers through casual observations or social interactions with employees. Working remotely likely severely reduces the amount of informal information managers have about their employees, and as a result, we posit that working remotely will influence employees' relative preferences for human versus AI-driven systems. However, we do not expect the impact to be uniform. We argue that those employees that feel that informal information about them is (not) to their advantage in the performance evaluation process will have a relatively (weaker) stronger preference for human managers when they work in the office, but this relatively (weaker) stronger preference for human managers will be reduced when working remotely.

Discrimination

One of the main problems caused by supervisors' tendency to include non-performance relevant informal data in their performance evaluation is discriminatory bias based on demographics (Kunda 1990; Biernat and Kobrynowicz 1997). While some informal data, such as an employee's favorite candy, may not have any negative stereotypes associated with it, other informal data, like skin color, frequently has. This is important because these negative stereotypes create lower expectations about employees' performance. Because of confirmation bias, these lower expectations result in managers focusing on (i.e., placing more weight on) those aspects of the formal performance data that are consistent with their lower expectations, which ultimately results in negatively biased assessments (Rabin and Schrag 1999; Frost et al 2015; Bol and Smith 2011). We predict that employees who have experienced workplace discrimination in the past, will have greater concerns that a manager will, consciously or unconsciously, be influenced by non-performance relevant informal information. These employees will also have fewer concerns about AI-driven systems' objectivity as they welcome being evaluated on just their performance. Thus, contrary to Newman et al. (2020), we argue that employees who believe they have suffered from discrimination in the past will have an appreciation for the AI-driven system's strict objectivity which will, all else equal, increase their preference for AI-driven systems relative to human-driven systems. This prediction is also consistent with the finding that members of underrepresented groups are more likely to select into and stay at firms that are already using AI-driven systems (Brown, Burke, and Sauciuc 2021).

We also predict that working remotely will moderate the impact of discrimination concerns. Specifically, we predict that working remotely will increase the relative preferences for AI by employees who have experienced workplace discrimination. The intergroup contact theory of prejudice posits that bias towards people in a different social group can be reduced by social interactions (Pettigrew 1998). In their meta-analysis of 515 research studies, Pettigrew et al. (2011), find that intergroup social contact is effective at increasing trust and forgiveness and reducing prejudice between groups. Consistent with this theory, we posit that because the social contact is reduced when working remotely, there is increased concern for discrimination by human supervisors. Employees who have been discriminated in the past are concerned that the lack of informal information about who they are, reduces them to just their stereotype (e.g., an employee becomes just a black lady to her manager, not Trenise who roots for the same sports team and comes in early like her manager to beat the traffic). Stronger stereotyping will likely result in more discriminatory bias. Thus, we hypothesize that the relative preference for AI-driven systems by employees who have experienced past discrimination will be stronger when employees work remotely than when they work in a shared workspace. We hypothesize a main effect and an interactive effect.

H2a: Employees who believe they have experienced workplace discrimination will show relatively higher preferences for AI-driven systems versus human-driven systems than employees who do not.

H2b: The effect of past discrimination on preferred evaluator type will be stronger when employees work remotely versus when they work in the office.

See Figure 1, Panel A for a graphical depiction of H2a and H2b.

Social intelligence

Some employees are particularly good at navigating social situations because they have high social intelligence. As a result, they can positively influence supervisors' expectations through informal conversations and informal information sharing. This is consistent with a long line of research that shows that positive social relationships and being an effective "political player" results in positively biased performance ratings (Prendergast and Topel 1996; Ittner, Larcker, and Meyer 2003; Bandiera et al. 2009). We predict that these employees will likely expect to benefit from managers' tendency to develop performance expectations using more than performance information and therefore have a relatively stronger preference for the subjectivity in human evaluation. On the other end of the spectrum, employees with low social intelligence are more likely to embrace AI as they find socialization tiresome and wish to avoid it. These employees prefer an AI that will not be influenced by friendships or political connections in its evaluation process. Again, we predict that working remotely will moderate these preferences. When working remotely, the opportunities to influence the supervisor are reduced, which again reduces the relative preference (or lack thereof) for human-driven performance evaluation. This leads to the following two hypotheses:

H2c: Employees who have lower social intelligence will show relatively higher preferences for AI-driven systems versus human-driven systems than employees who have higher social intelligence.

H2d: The effect of social intelligence on preferred evaluator type will be reduced when employees work remotely versus when they work in the office.

See Figure 2, Panel A for a graphical depiction of H2c and H2d.

IV. METHODS

We test our hypotheses using two experiments. Both were conducted on Amazon's Mechanical Turk (AMT) platform and were approved by the IRB of the major US-based research institution where data was collected. Prior research has shown that this platform supplies participants who behave similarly to traditional student-driven samples, while also being more demographically representative of the American labor force (Paolacci, Chandler, Ipeirotis 2010; Farrell, Grenier, and Leiby 2017; Buchheit et al. 2018). A representative sample of the broader labor force is important for our research question, in which we explore the preferences and perceptions of rank-and-file employees at firms. In addition, AMT participants typically have experience working both online and in person meaning that they have received digital performance reviews for their AMT work done in a remote setting with no social interaction and traditional human-driven evaluations from bosses in shared office space with social interactions. This set of dual work experiences allows them to be well suited to imagining a world where both types of reviews are possible and to be able to express a preference between them. Participants earned a base pay of $2.00 and then an additional $0.10 for each of two attention check questions and one manipulation check question they answered correctly. Participants were told the task would take up to 20 minutes. All participants were adults who were at least 18 years old.

Several steps were taken to increase the quality of the data collected from AMT. The experiments were conducted using Cloud Research, an independent research platform which allows for more rigorous subject filtering. The participant pool was limited to workers who had completed at least 1,000 tasks with a 90% approval rating and whose IP addresses were located in the United States. Participants were also excluded if they had duplicate IP addresses or had IP addresses identified by Cloud Research as "suspicious." Participants first completed a consent form and then a captcha, a picture of text that participants needed to translate into machine readable data. Participants who could not successfully complete the captcha were not allowed to continue. The use of a captcha is in keeping with best practices surrounding AMT usage and helps filter for automated, non-human, participants.

Both experiments use a similar design and company setting. They are 1x2 experiments with additional measured variables. In each, participants are asked to assume the role of a salesperson at ABC Robotics. They learn that they spend half their time selling robots and the other half on site with clients conducting trainings. They are evaluated on both tasks using objective, and subjective performance metrics to determine whether they will receive a monthly performance-based bonus. See Figure 3 for the detailed information participants saw regarding their evaluation. They are told that this bonus is meaningful to them, and that in the months that they receive it, they use it to buy themselves a special treat.

After their job has been explained to them, they are told that ABC Robotics is forming a new division to which they will be reassigned. At the moment, ABC Robotics is using human-driven evaluation for some departments and using AI-driven evaluation for others and participants are told that these systems have been equally successful. The company needs to determine which system to use in the new division and are therefore conducting a vote to get employees' opinions. Participants are asked to vote on which they would prefer. Their votes are measured on a 1-5 scale where 1 is "strongly prefer humans," 3 is no preference, and 5 is "strongly prefer AI." After participants vote, they complete a post-experiment questionnaire (PEQ). The PEQ contains demographic data and the measured variables used in analysis and tests of hypotheses.

While the base of the experiments is the same between the two experiments, as discussed above, they each have a different manipulation. The PEQ also differs slightly between the two experiments. The next two sections detail these differences. Figure 3 displays the information all participants were given about their performance data. Figure 4 provides a timeline which highlights similarities and differences.

Experiment 1 Design

In experiment 1, we manipulate the *Environment Stability*. The company and the job are as described in the previous section, but the environment of the firm is manipulated between subjects at two levels: *Stable* and *Unstable*. In the *Stable* condition, participants learn that the firm's current operating environment is not different from prior years, and the company has extensive historical employee performance data and experience conducting performance evaluations under the current economic conditions. In the *Unstable* condition, participants learned that the COVID-19 pandemic and related financial crisis have significantly disrupted the company's operations and the firm does not have experience evaluating employee performance under the current operating conditions. See Figure 5 for the exact manipulation. All participants are told that they work from home when they are selling the robots and in their client's offices when they are training. As a manipulation check, participants were asked which economic condition they are in. Participants who cannot answer this question correctly are removed from the sample for data analysis.

In a PEQ, participants reported whether they felt they had been subject to workplace discrimination during their career (outside the experiment). Participants were asked to rate their agreement with the statement, "I have been subject to discrimination at work." Responses were recorded on a seven-point scale from "strongly disagree" (coded as 1) to "strongly agree" (coded as 7). These responses constitute the *Past Discrimination* variable in the analysis. Demographic information and other perceptions relating to AI-driven and human-driven performance evaluations were also collected in the PEQ.

Experiment 2 Design

As stated before, experiment 2 uses the same company setting, job, and dependent variable (*Preferred Evaluator*) as experiment 1. All participants are told that they are in a stable environment. In experiment 2, we manipulate working remotely and name the variable *Workspace*. *Workspace* is manipulated between subjects at two levels: *Shared* and *Remote*. In the *Shared* condition, participants were informed that they work in a shared central office whenever they are not on-site with customers. In the *Remote* condition, participants were told that they work from home whenever they are not on-site with customers. To increase the saliency of this manipulation, participants are also shown a photo of a desk on a white background and are told to imagine it is their desk. They are also given the same details about features of their offices. For example, they are told that they have snacks and coffee available to them. See Figure 6 for the exact manipulation. After participants see the *Workspace* manipulation, they are asked to report where they work for their job at ABC robots. Participants are not allowed to advance in the experiment until they correctly respond to this manipulation check.

After the manipulation check, participants indicate their preference for an AI-driven or human-driven evaluation system and then complete a PEQ. The PEQ is similar to the one of experiment 1, but the Tromsø Social Intelligence Scale is added. This scale was developed and validated in Silvera, Martinussen and Dahl (2001). Participants responded to 21 items using a seven-point scale ranging from “strongly disagree” (coded as 1) to “strongly agree” (coded as 7). The questions include statements like “I understand other peoples' feelings” and “I find people unpredictable.” See the appendix for the complete scale

V. RESULTS

A total of 150 and 155 participants were recruited from AMT for experiments 1 and 2, respectively. Participants had a mean age of 38 years (36 years in experiment 2) with a reported mean work experience of 16 years (12 years in experiment 2). For sample descriptive statistics, see Table 1 for experiment 1 and Table 2 for experiment 2. All participants completed manipulation checks. In experiment 1, 124 participants (83%) correctly responded to the manipulation check. These participants are retained for analysis in experiment 1. In experiment 2, all 155 participants passed the manipulation check because participants could not move forward in the task without correctly answering the question and were referred back to the pertinent information for additional review if they submitted an incorrect response.

Tests of H1

In the full sample of experiment 1, the mean value of *Preferred Evaluator* is 2.73 out of 5 where 3 is the midpoint of the scale, indicating that participants slightly prefer human evaluators over AI on average. However, consistent with H1 which predicts that employees will show a stronger relative preference for AI-driven systems when the operating environment is stable than when it is unstable, mean *Preferred Evaluator* in the *Stable* condition is 2.93 and 2.43 in the *Unstable* condition. See Table 3, Panel A for simple means by condition. To more formally test for the effect of *Environment Stability* on *Preferred Evaluator*, we perform an analysis of covariance including *Environment Stability* and *Past Discrimination* (Table 3, Panel B). We find a significant effect of *Environment Stability* on *Preferred Evaluator* ($F_{(1,121)} = 11.38$; $p < 0.01$), consistent with H1.[2]

[2] In an untabulated result, we repeat the analysis removing *Past Discrimination* from the model and continue to find a significant effect of *Environment Stability* on *Preferred Evaluator* ($p = 0.01$).

To further investigate the mechanism by which *Environment Stability* affects *Preferred Evaluator*, we perform a mediation analysis using the simultaneous regression method outlined in Hayes (2018). In development of H1, we argue that employees believe that an AI-driven system is less able than a human-driven system to subjectively adjust performance for the relevant performance context when the operating environment is unstable, resulting in performance evaluations that are perceived to be less fair. We therefore test whether the perceived ability of AI and human managers to consider the context of the operating environment mediates the effect of *Environment Stability* on *Preferred Evaluator*. To capture concerns about ability to consider context, we use participants' responses to two post-experimental questionnaire items: (1) "When casting my vote for who would evaluate me (AI or a human manager), I was concerned that the AI wouldn't be able to fairly consider the circumstances I was in." (2) "When casting my vote for who would evaluate me (AI or a human manager), I was concerned that a human manager wouldn't be able to fairly consider the circumstances I was in." These responses are recorded on a seven-point scale from "strongly disagree" (coded as 1), to "strongly agree" (coded as 7). Responses to the first of these PEQ items are captured in the *AI Context* variable, and responses to the second item are captured in the *Human Context* variable. Responses to both items are included in the mediation analysis, and *Past Discrimination* is included as a covariate.

Full results of the mediation analysis are presented in Table 4 and depicted graphically in Figure 7. We find evidence that *Environment Stability* affected participants' concerns about AI's ability to consider the relevant context ($p < 0.01$, see Table 4, Panel A), but we do not find evidence of a significant effect of *Environment Stability* on concerns about a human manager's ability to fairly consider the context of employee performance ($p = 0.12$, see Table 4, Panel B). We also find that each of these measures significantly predicted *Preferred Evaluator* ($p < 0.01$ for both *AI*

Context and *Human Context*, see Table 4, Panel C). We find evidence that the total indirect effect of *Environment Stability* on *Preferred Evaluator* through *AI Context* is significant (95% confidence interval: [0.04, 0.44], see Table 4 Panel D). We do not find evidence of a similar indirect effect through *Human Context* (95% confidence interval: [-0.04, 0.40]). These results indicate that participants preferred an AI-driven evaluation system more in a stable environment than an unstable environment because they were less concerned about AI's objectivity in a stable environment. Consistent with the underlying reasoning for our hypothesis, environment stability did not affect participants' concern about a human manager's ability to contextualize performance information. Overall, we find strong support for H1.

Tests of H2a and H2b

H2a predicts that employees who believe they have been discriminated against in the past will show relatively higher preferences for AI-driven systems versus human-driven systems than employees who do not. Recall that we measure perceived past discrimination by capturing participants' agreement with the following statement, "I have been subject to discrimination at work." (see Table 1, Panel A). Participants in experiment 1 rated their agreement as a mean of 3.44 out of 7 for experiencing discrimination in the past. Thirty-two percent of participants reported that they at least "somewhat agree" (a response of five or more) with the statement. We check that randomizing our participants into conditions was successful and find that *Past Discrimination* is not significantly different between *Stable* and *Unstable* conditions. See Table 5 Panel A for more details. We formally test H2a with the model presented in Table 3. We find that *Past Discrimination* significantly predicts *Preferred Evaluator* ($p < 0.01$), consistent with H2a.

Additional tests of H2a and H2b are conducted using data collected in experiment 2. We test the effect of *Past Discrimination, Workspace,* and their interaction on *Preferred Evaluator* using an ANCOVA model.[3] We again find that *Past Discrimination* significantly predicts *Preferred Evaluator* ($p < 0.01$; Table 6, Panel A).[4] Our study thus finds robust support for H2a using two different experiments with two different experimental samples. H2b predicts that the effect of past discrimination on preferred evaluator type will be larger when employees have limited opportunities for social contact compared to when they have ample opportunities for social contact. We find an interactive effect between *Workspace* and *Past Discrimination* such that past discrimination has a stronger effect on preferred evaluator when working remotely than when in a shared office ($p = 0.03$). Further analysis (Table 6, Panel A) shows that the effect of *Past Discrimination* on *Preferred Evaluator* is significant ($p < 0.01$) when *Workspace* is remote, and not significant in a shared office ($p = 0.23$).[5]

To determine if the interaction is consistent with the predicted pattern, we separate participants into four groups using a median split of *Past Discrimination* and *Workspace.* We then examine the mean *Preferred Evaluator* in each of these groups (see Table 6, Panel B). This pattern is consistent with our prediction. We find that participants who have experienced more workplace discrimination have a stronger preference for AI-driven evaluation systems than participants who have experienced less workplace discrimination in both conditions of *Workspace.* Further, as predicted in H2b, we find that participants who have experienced more past discrimination have a stronger preference for an AI-driven performance evaluation when working remotely than when working in a shared office.

[3] In untabulated results, we run a similar analysis while including the interaction between *Social Intelligence* and *Workspace* as an additional independent variable. Inferences made from both models are consistent. In fact, results of this alternate model are generally stronger.

[4] We once again check that *Past Discrimination* is not significantly different between the two *Workspace* conditions (Table 5, Panel B).

[5] Additionally, in untabulated analysis when examining under- and over-represented demographics as the independent variable instead of self-reported past discrimination, similar inferences are found.

Tests of H2c and H2d

H2c predicts that employees who have lower social intelligence will show relatively higher preferences for AI-driven systems versus human-driven systems than employees who have higher social intelligence. H2d predicts that the effect of social intelligence on preferred evaluator type will be smaller when employees work remotely than when they work in a shared office . The minimum *Social Intelligence* in our sample is 51, the maximum is 145, the mean is 94.28 and the standard deviation is 17.96 (see Table 2). To ensure there is balance between the cells, we test and find that *Social Intelligence* is not significantly different between the two conditions (see Table 5, Panel B). We conduct a median split on *Social Intelligence* and find participants with low social intelligence have a stronger relative preference for an AI-driven system (mean *Preferred Evaluator* of 3.38) than participants with high social intelligence (mean *Preferred Evaluator* of 2.46).

We use an ANCOVA model to test the effect of *Social Intelligence*, *Workspace*, and their interactive effect on *Preferred Evaluator* while controlling for *Past Discrimination* (see Table 7, Panel A). We find a significant main effect of *Social Intelligence* ($F_{(1,150)} = 4.07$, $p = 0.04$), and a marginally significant interactive effect of *Social Intelligence* and *Workspace* ($F_{(1,150)} = 3.53$, $p = 0.06$). Consistent with H2d, we find that the effect of *Social Intelligence* on *Preferred Evaluator* is stronger in the *Shared* condition than in the *Remote* condition. These results support H2c and H2d.

Additional Analysis

We posit in the theory section that employee preferences are based on their perceptions of when a performance evaluation would be fairer. However, it is possible that instead of fairness, participants value some other feature of the two systems. For example, they may believe that the human will be a more lenient judge of performance or that the AI will be easier to deceive, resulting in greater likelihood of a reward. In order to confirm that participants are selecting the evaluator based on fairness concerns, we also directly asked participants which evaluator they believed would be the fairest. *Fairest Evaluator* is measured using a scale from 1 to 5 where 1 is a human and 5 is an AI.

We first examine the univariate relationship between *Fairest Evaluator* and *Preferred Evaluator*. The two are significantly positively correlated with one another in both experimental populations ($p < 0.01$) (Table 1 Panel B, and Table 2, Panel B). This raw correlation provides initial evidence supporting our theory. We then re-examine our hypotheses by performing similar tests but replacing *Preferred Evaluator* with *Fairest Evaluator* as the dependent variable. Consistent with H1, we find that *Environment Stability* significantly predicts *Fairest Evaluator* ($p = 0.02$) (Table 8, Panel B) where AI is relatively more preferred in a stable environment. We also find a significant effect of *Past Discrimination* on *Fairest Evaluator* with the data from experiment 1 ($p < 0.01$, Table 8, Panel B), consistent with H2a. People who report having been discriminated against are more likely to believe that AI-driven systems will be fairer than those who did not report that. In Table 9, we test whether *Past Discrimination* and *Workspace* have an interactive effect on *Fairest Evaluator.* Although we find support for H2b in our main analysis, we do not find a significant result ($p = 0.97$) in this supplemental analysis. It is unclear whether there is a theoretical reason for this null result or whether it is driven by limitations of our study design. We leave the examination of this to future research. As shown in Table 10, Panel A, we do not find a significant main effect of *Social Intelligence* on *Fairest Evaluator*. We do, however, find a significant interactive effect ($F_{(1, 150)} = 3.62$, $p = 0.06$) of *Social Intelligence* and *Workspace* on *Fairest Evaluator*. We also demonstrate through simple effects tests that the effect of *Social*

Intelligence on *Fairest Evaluator* is significant (p = 0.02) in the *Shared* condition, and not significant in the *Remote* condition (p = 0.79). These results are consistent with H2d.

We continue to explore the role of fairness by testing whether *Fairest Evaluator* mediates the effect of *Social Intelligence* on *Preferred Evaluator* and whether this indirect effect is moderated by *Workspace* (see Figure 8 for visual depiction). To test this moderated mediation model, we use the simultaneous OLS regression method and Model 7 of the PROCESS macro as outlined in Hayes (2018). This method allows us to simultaneously test whether *Fairest Evaluator* mediates the effect of *Social Intelligence* on *Preferred Evaluator* and whether such an effect is moderated by *Workspace.* In doing so, we connect our prior analyses on how the independent variables of interest affect participants' perceptions of the fairest evaluator and, in turn, how perceptions of fairness affect their preferred evaluator type. Regressions are conducted using 5,000 bootstrap samples and a 95% confidence interval. We find a significant interactive effect of *Social Intelligence* and *Workspace* on *Fairest Evaluator* (t = 1.90, p = 0.03, one-tailed test) (see Table 11). Additionally, we find that *Fairest Evaluator* significantly predicts *Preferred Evaluator* (t = 6.18, p < 0.01). We also find evidence of a statistically significant indirect effect of *Social Intelligence* on *Preferred Evaluator* through *Fairest Evaluator* in the *Shared* condition (95% confidence interval: [-0.0160, -0.0007]) but not in the Remote condition (95% confidence interval: [-0.0062, 0.0091]). The index of moderated mediation tests for a significant difference in the strength of the indirect effect between levels of *Workspace* and is nearly significant at a 95% confidence level (confidence interval: [-0.0003, 0.0208]). These results indicate that in an environment where there are ample opportunities for social contact, people with higher social intelligence are less likely to prefer an AI-driven evaluation system than people with lower social intelligence because they believe AI is less capable of making a fair evaluation than a human manager. In sum, our results not only support the hypotheses, but also the underlying reasoning.

VI. CONCLUSION

The COVID-19 pandemic increased the economic environment's overall instability and made working remotely a mainstream work option. At the same time, AI technology emerged as a viable alternative to human-driven performance evaluation systems. These changes triggered many organizations to rethink their employee performance assessment systems, specifically to assess whether their human-driven system should be replaced by an AI-driven system. Managers seem to be excited about the potential of AI-driven systems, but it is still unclear how successful the systems will be at motivating employees. Success is difficult to predict because it is not only dependent on how technologically advanced the systems are, but also, on how employees will react to having their performance evaluated by machines instead of by humans.

While previous research that examines employee preferences finds that employees always prefer human evaluators, we predict and find that preferences between AI and human evaluators are not uniform. Instead, they are driven by fairness concerns and employees' relative preferences for objective measurement versus subjective, judgement-based evaluations. Specifically, we find that in more stable economic environments, employees show a relatively higher preference for AI-driven performance evaluation systems. Employees who have been discriminated against also show a relatively higher preference for AI than those who have not, and this effect is stronger in remote settings than in shared offices. Moreover, we find social intelligence predicts preferences for AI, but only when employees are working in a shared office space where there are ample opportunities for social contact. Finally, we show that these changes in preference occur because situational factors affect employees' beliefs about which evaluator will be the fairest. Firms can use these findings to guide investments into AI-driven evaluation systems by better understanding in which settings employees are more likely to accept the systems as fair. For example, our study suggests that companies in mature industries with more predictable operating environments or with historical struggles with workplace discrimination may be better candidates for AI-driven performance evaluation.

While this paper is an important first step to examine cross-sectional differences in employees' preferences towards AI, it only examines a few situations. We hope that future research will continue to examine other settings and add to these results. Moreover, our results rest on employees' *perceptions* of AI-driven performance evaluation because most people are currently not being evaluated by a purely AI-driven system. How these perceptions change when more explanation is provided about how the technology works will be an important addition to this line of research. Moreover, since different AI-driven evaluation systems function differently, it would also be valuable to examine how preferences differ depending on the specific (technological) features of the AI.

Our study is an important first step in management accounting research investigating the effects of AI on performance evaluation. This new technology is being used by a small but increasing number of firms. It is important that management accountants join in the discussion of how and when to implement it in firms. This discipline's deep understanding of the evaluation process can help developers create more effective technology. Moreover, it can help managers understand which employee groups may be a better fit for adoption of an AI-driven system and communicate to employees how these systems will differ from traditional human-driven systems.

BIBLIOGRAPHY

Baker, S. R., Bloom, N., Davis, S. J., & Terry, S. J. (2020). Covid-induced economic uncertainty (No. w26983). *National Bureau of Economic Research*.

Baker, G., Gibbons, R., & Murphy, K. J. (1994). Subjective performance measures in optimal incentive contracts. *The quarterly journal of economics*, 109(4), 1125-1156.

Bandiera, O., Barankay, I., & Rasul, I. (2009). Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4), 1047-1094.

Berger, B., Adam, M., Rühr A., & Benlian, A. (2021). Watch me improve – Algorithm aversion and demonstrating the ability to learn. *Business & Information Systems Engineering*, 63.1: 55-68.

Biernat, M., and T. K. Vescio. 2002. She Swings, She Hits, She's Great, She's Benched: Implications of Gender-Based Shifting Standards for Judgment and Behavior. *Personality and Social Psychology Bulletin*, 28(1), 66–77.

Bol, J. C. (2008). Subjectivity in Compensation Contracting. *Journal of Accounting Literature*, 27, 1-24.

Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, 86(5), 1549-1575.

Bol, J. C., & Leiby, J. (2018). Subjectivity in Professionals' Incentive Systems: Differences between Promotion-and Performance-Based Assessments. *Contemporary Accounting Research*, 35(1), 31-57.

Bol, J. C., Margolin M., Schaupp D., (2021). Multi-Rater Performance Evaluation and Calibration: Managing Multiple Opinions. Working paper.

Brown, J., Burke, J., & Sauciuc, A. (2021). *Workforce Diversity and Artificial Intelligence: Implications for AI Integration into Performance Evaluation Systems*. Working Paper.

Buchheit, S., Doxey, M. M., Pollard, T., & Stinson, S. R. (2018). A technical guide to using Amazon's Mechanical Turk in behavioral accounting research. *Behavioral Research in Accounting*, 30(1), 111-122.

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.

Cappelli, P., & Tavis, A. (2016). The performance management revolution. Harvard Business Review. https://hbr.org/2016/10/the-performance-management-revolution

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809-825.

Chan, H & Dimauro, J. (2020). Black Lives Matter movement sparks outcry for corporations to show diversity gains. Retrieved from https://blogs.thomsonreuters.com/answerson/black-lives-matter-corporate-diversity-gains/

Cheng, M. M., & Hackett, R. D. (2021). A critical review of algorithms in HRM: Definition, theory, and practice. *Human Resource Management Review*, 31(1), 100698.

Demeré, B. W., Sedatole, K. L., & Woods, A. (2019). The role of calibration committees in subjective performance evaluation systems. *Management Science*, 65(4), 1562-1585.

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.

Downes, P. E., & Choi, D. (2014). Employee reactions to pay dispersion: A typology of existing research. *Human Resource Management Review*, 24(1), 53-66.

Du, F., Tang, G., & Young, S. M. (2012). Influence activities and favoritism in subjective performance evaluation: Evidence from Chinese state-owned enterprises. *The Accounting Review*, *87*(5), 1555-1588.

Engellandt, A., & Riphahn, R. T. (2011). Evidence on incentive effects of subjective performance evaluations. *ILR Review*, 64(2), 241-257.

Farrell, A. M., Grenier, J. H., & Leiby, J. (2017). Scoundrels or stars? Theory and evidence on the quality of workers in online labor markets. *The Accounting Review*, 92(1), 93-114.

Fecheyr-Lippens, B., Schaninger, B., & Tanner, K. (2015). Power to the new people analytics. McKinsey Quarterly. https://www.mckinsey.com/business-functions/organization/our-insights/power-to-the-new-people-analytics#

Fisher, J. G., Maines, L. A., Peffer, S. A., & Sprinkle, G. B. (2005). An experimental investigation of employer discretion in employee performance evaluation and compensation. *The Accounting Review*, 80(2), 563-583.

Frost, P., B. Casey, K. Griffin, L. Raymundo, C. Farrell, and R. Carrigan. 2015. The influence of confirmation bias on memory and source monitoring. *The Journal of General Psychology*, 142(4), 238–252.

Gibbs, M. J., Merchant, K. A., Stede, W. A. V. D., & Vargus, M. E. (2005). The benefits of evaluating performance subjectively. *Performance Improvement*, 44(5), 26-32.

Glassdoor. (2019). Diversity & Inclusion Study 2019. Retrieved from https://about-content.glassdoor.com//app/uploads/sites/2/2019/10/Glassdoor-Diversity-Survey-Supplement-1.pdf.

Greene, T. (2018, July 10). IBM is using its AI to predict how employees will perform. Retrieved from https://thenextweb.com/artificial-intelligence/2018/07/10/ibm-is-using-its-ai-to-predict-how-employees-will-perform/.

Hayes, A. F. (2018). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford publications.

Hu, W. (2021). The Lost Productivity: An Experimental Investigation of Human Versus Algorithm-Based Discretion in Incomplete Compensation Contracts. Working paper.

Holsinger, L., et al. (2019). Performance Transformation in the Future of Work. Mercer. https://www.mercer.com/our-thinking/career/performance-transformation-in-the-future-of-work.html

Ittner, C. D., Larcker, D. F., & Meyer, M. W. (2003). Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review*, 78(3), 725-758.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of labor Economics*, 26(1), 101-136.

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.

Knight, R. (2020). How to Do Performance Reviews — Remotely. Harvard Business Review. https://hbr.org/2020/06/how-to-do-performance-reviews-remotely

Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, 108(3), 480–498.

Kuvaas, B. (2006). Performance appraisal satisfaction and employee outcomes: mediating and moderating roles of work motivation. *The International Journal of Human Resource Management*, *17*(3), 504-522.

Lund, S. Madgavkar, A., Manyika, J., Smit, S., Ellingrud, K., and Robinson, O. (2021). The future of work after COVID-19. McKinsey Global Institute. https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-work-after-covid-19

Mackenzie, L. N., Wehner, J., & Kennedy, S. (2020). How do you evaluate performance during a pandemic? Harvard Business Review. https://hbr.org/2020/12/how-do-you-evaluate-performance-during-a-pandemic

Masterson, S. S., Lewis, K., Goldman, B. M., & Taylor, M. S. (2000). Integrating justice and social exchange: The differing effects of fair procedures and treatment on work relationships. *Academy of Management journal*, *43*(4), 738-748.

Moise, I. & Cruise, S. (2020). Goldman Sachs executive's email making plea for racial equality goes viral at firm. Retrieved from https://www.reuters.com/article/us-usa-goldman-sachs-race/goldman-sachs-executives-email-making-plea-for-racial-equality-goes-viral-at-firm-idUSKBN23C086

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*, 160, 149-167.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5), 411-419.

Pettigrew, T. F. (1998). Intergroup contact theory. Annual review of psychology, 49(1), 65-85.

Pettigrew, T. F., Tropp, L. R., Wagner, U., & Christ, O. (2011). Recent advances in intergroup contact theory. *International journal of intercultural relations*, *35*(3), 271-280.

Poon, J. M. (2004). Effects of performance appraisal politics on job satisfaction and turnover intention. *Personnel Review.*

Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted?. *Journal of Forecasting*, *36*(6), 691-702.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 37(1), 7-63.

Prendergast, C., & Topel, R. (1993). Discretion and bias in performance evaluation. *European Economic Review*, 37(2-3), 355-365.

Prendergast, C., & Topel, R. (1996). Favoritism in organizations. *Journal of Political Economy*, 104(5), 958-978.

Rabin, M., and J. L. Schrag. 1999. First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics*, 114(1), 37–82.

Rogel, C. (2020). How much do performance reviews actually cost and are they really worth it? Decisionwise. https://decision-wise.com/how-much-do-performance-reviews-actually-cost-and-are-they-really-worth-it/

Rosenbaum, E. (2019, April 3). IBM artificial intelligence can predict with 95% accuracy which workers are about to quit their jobs. Retrieved from https://www.cnbc.com/2019/04/03/ibm-ai-can-predict-with-95-percent-accuracy-which-employees-will-quit.html

Silvera, D., Martinussen, M., & Dahl, T. I. (2001). The Tromsø Social Intelligence Scale, a self-report measure of social intelligence. Scandinavian journal of psychology, 42(4), 313-319.

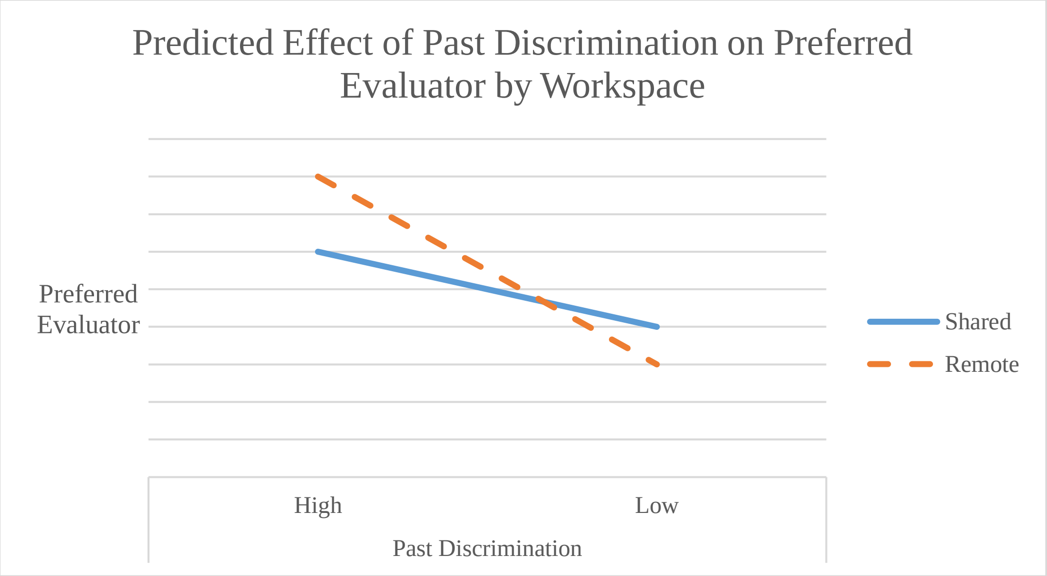
Simon, H. A. (1990). Bounded rationality. In Utility and probability (pp. 15-18). Palgrave Macmillan, London.

Simons, T., & Roberson, Q. (2003). Why managers should care about fairness: the effects of aggregate justice perceptions on organizational outcomes. *Journal of Applied Psychology*, 88(3), 432.

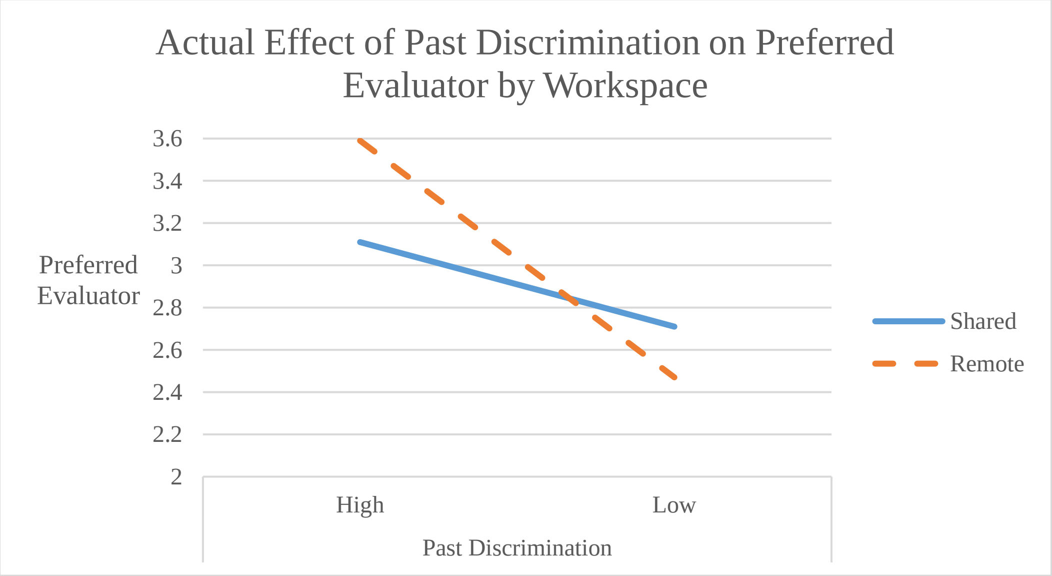
Trevor, C. O., Reilly, G., & Gerhart, B. (2012). Reconsidering pay dispersion's effect on the performance of interdependent work: Reconciling sorting and pay inequality. *Academy of Management Journal*, 55(3), 585-610.

Figure 1: Predicted and Actual Results for H2a and H2b

Panel A: Predicted Results



Panel B: Actual Results

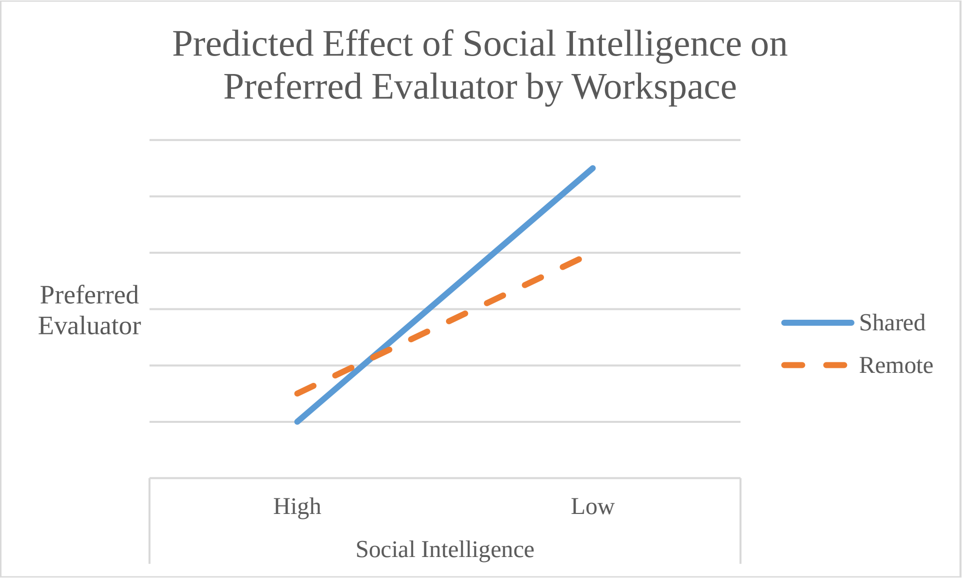


Panel A documents the hypothesized pattern of *Preferred Evaluator* by *Past Discrimination* and *Workspace* where lower values of *Preferred Evaluator* indicate a stronger preference for human evaluators and higher values indicate a stronger preference for AI.

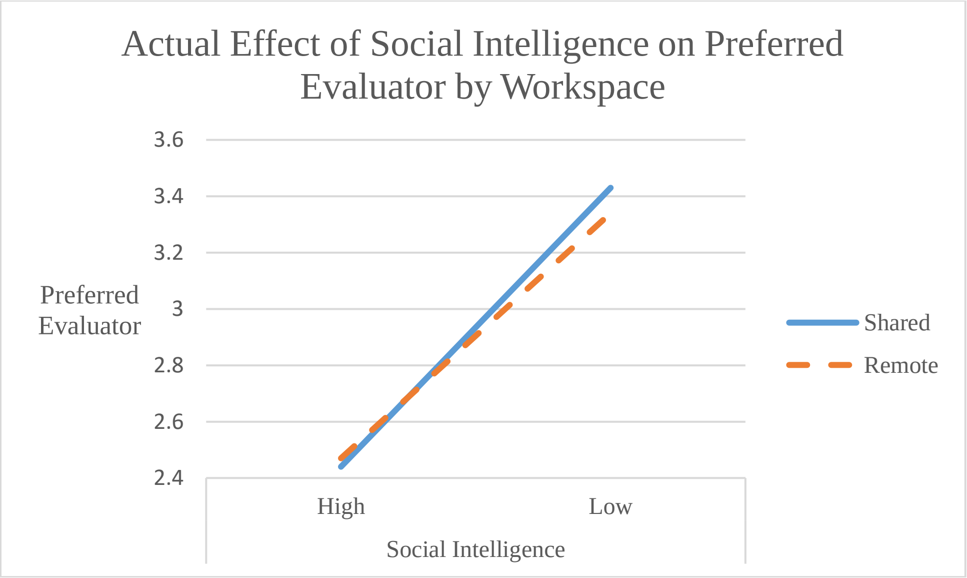
Panel B documents the values of *Preferred Evaluator* for participants who are either above or below the median value of *Past Discrimination* by *Workspace* condition. This data is from experiment 1 and is shown in greater detail in Table 6, Panel B. See Table 1 for variable definitions.

Figure 2: Predicted and Actual Results for H2c and H2d

Panel A: Predicted Results



Panel B: Actual Results



Panel A documents the hypothesized pattern of *Preferred Evaluator* by *Social Intelligence* and *Workspace* where lower values of *Preferred Evaluator* indicate a stronger preference for human evaluators and higher values indicate a stronger preference for AI.

Panel B documents the values of *Social Intelligence* for participants who are either above or below the median value of *Past Discrimination* by *Workspace* condition. This data is from experiment 2 and is shown in greater detail in Table 7, Panel B. See Table 1 for variable definitions.

Figure 3: Performance Data Format Information

Participants in both experiments saw the following information:

To assess your sales performance, the company records your total sales, the prices you negotiated, and the premium add-ons that you sold. In addition, to help put your sales numbers into perspective and adjust for factors outside your control, the company collects several other pieces of information. For example, the company collects national and local economic indicators that are relevant to your sales performance during the month such as global oil prices, local unemployment levels, federal interest rates, and extreme weather. The company also collects information on competitors like their market share and pricing strategy. At the end of the month, your sales performance is evaluated by analyzing your sales number with consideration for these context variables described above.

To assess your performance as a trainer, customers fill out a survey about their satisfaction with the training. Some of the survey questions are numerical and answered on a standardized scale, while others are open-ended and allow the customer to provide a free response. Open-ended questions are just as important as the numerical questions because they provide more specific information about your strengths and weaknesses as a trainer. The organization also collects information that could affect customer satisfaction with the training such as trends in robot usage and the prior experience the customers you trained had with robotics. As with sales data, the training performance is evaluated by analyzing the survey responses in light of the customer types that you served.

Your performance is weighted so that 50% of your final evaluation is based on sales performance, and 50% is based on your effectiveness as a trainer.

Figure 4: Timeline of Experiments

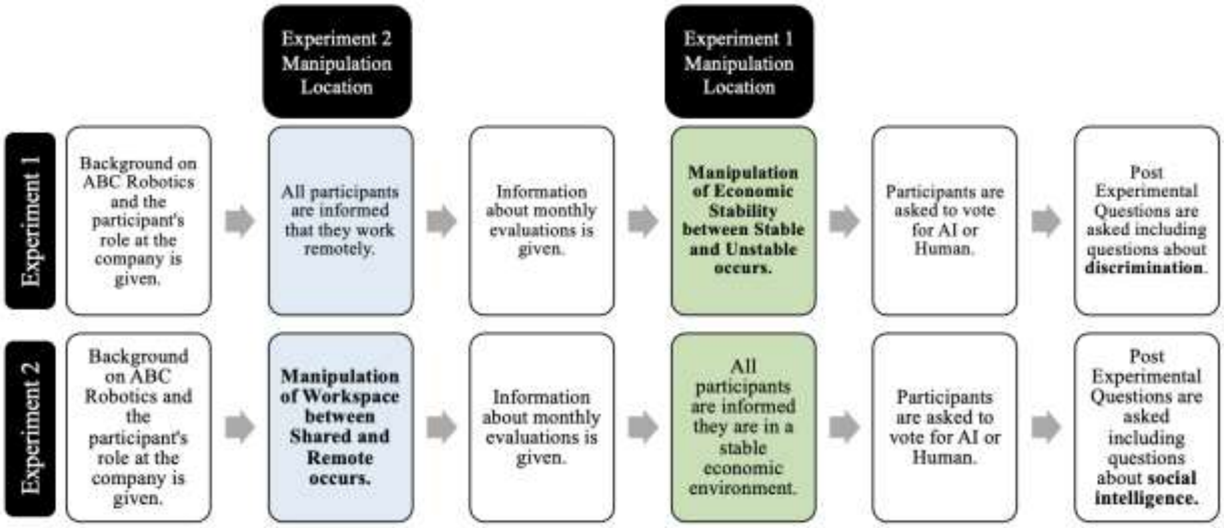

Experiment 2 Manipulation Location
Experiment 1 Manipulation Location
Experiment 1
Background on ABC Robotics and the participant's role at the company is given.
All participants are informed that they work remotely.
Information about monthly evaluations is given.
Manipulation of Economic Stability between Stable and Unstable occurs.
Participants are asked to vote for AI or Human.
Post Experimental Questions are asked including questions about discrimination.
Experiment 2
Background on ABC Robotics and the participant's role at the company is given.
Manipulation of Workspace between Shared and Remote occurs.
Information about monthly evaluations is given.
All participants are informed they are in a stable economic environment.
Participants are asked to vote for AI or Human.
Post Experimental Questions are asked including questions about social intelligence.

Figure 5: Environment Stability Manipulation

Stable:

The quality of the data at ABC Robotics is high. There is historical data on all variables for several years and the data set is complete. The current business environment mostly stable. Both the AI algorithm and the human manager have experience assessing performance under these circumstances.

Unstable:

The quality of the data at ABC Robotics is high. There is historical data on all variables for several years and the data set is complete. The current business environment, however, is highly unstable: the world has been hit by the COVID-19 pandemic and related financial crisis. Neither the AI algorithm nor the human manager has any experience with assessing performance under these circumstances.

Figure 6: Opportunity for Social Contact in the Workspace Manipulation

Shared Office:

You work as a salesperson selling the robots and conducting on-site trainings with employees at the companies that buy the robots. At ABC Robotics, everyone, including senior leadership, works in a shared office space whenever they are not with their clients. You spend about half of your time with clients and the other half of your time working from the office finding sales leads, doing paperwork, etc.

ABC Robots has a comfortable and well decorated office space. You have a large desk with a top quality office chair, multiple computer monitors, and lots of healthy snacks for you to eat throughout the day. You even have a fancy coffee machine for your morning pick me up.

Below you will see a picture of the desk you work at.

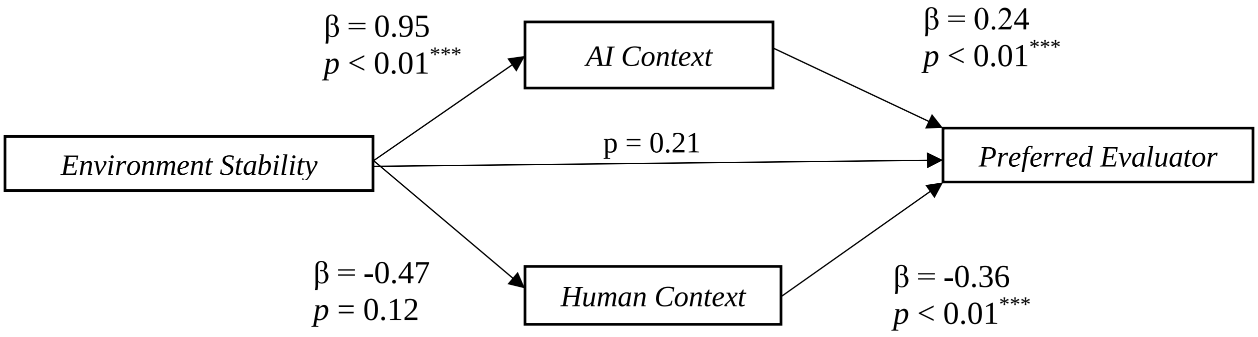
Remote:

You work as a salesperson selling the robots and conducting on-site trainings with employees at the companies that buy the robots. At ABC Robotics, everyone, including senior leadership, works from home whenever they are not with their clients. You spend about half of your time with clients and the other half of your time working from home finding sales leads, doing paperwork, etc.

You have a comfortable and well decorated home office space. You have a large desk with a top quality office chair, multiple computer monitors, and lots of healthy snacks for you to eat throughout the day. You even have a fancy coffee machine for your morning pick me up.

Below you will see a picture of the desk you work at.

Figure 7: Mediation of the Effect of *Environment Stability* on *Preferred Evaluator*

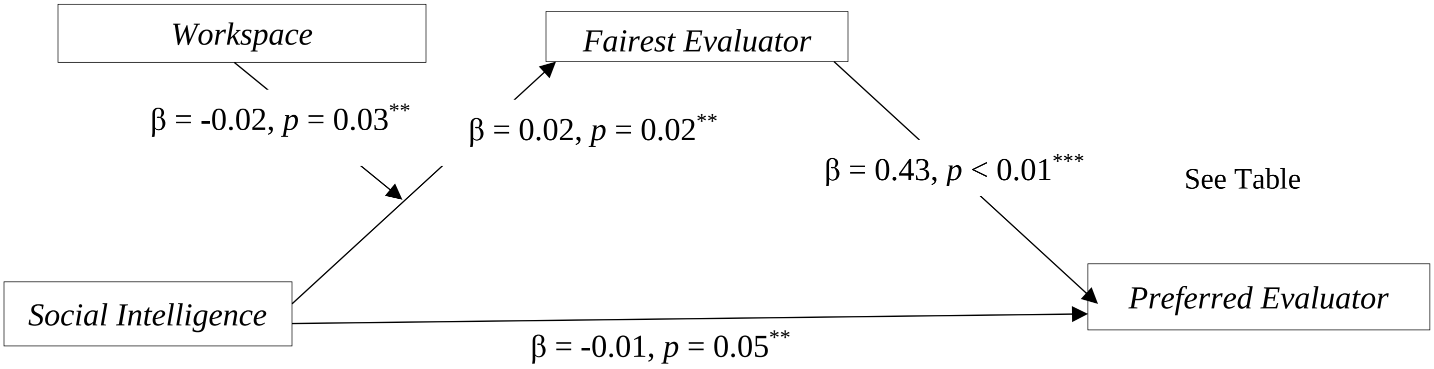


Indirect effects of *Environment Stability* on *Preferred Evaluator*

	Lower CI	Upper CI
through *AI Context*	0.04	0.44
through *Human Context*	-0.04	0.40

See Table 4 for detailed explanation of the regression analysis.

Figure 8: The Effect of *Social Intelligence* on *Preferred Evaluator* through *Fairest Evaluator*, Moderated by *Workspace*



Indirect effects of *Social Intelligence* on *Preferred Evaluator* through *Fairest Evaluator* for each level of *Workspace*

	Lower CI	Upper CI
Shared	-0.0160	-0.0007
Remote	-0.0062	0.0091

See Table 11 for detailed explanation of the regression analysis.

Table 1: Experiment 1 Descriptive Statistics and Pearson Correlation Table

Panel A: Descriptive Statistics

Parameter	*Min*	*Max*	*Median*	*Mean*	*SD*
Preferred Evaluator	1	5	3	2.73	1.30
Fairest Evaluator	1	5	3	2.76	1.15
Past Discrimination	1	7	3	3.44	1.89
Age (years)	18	73	35	38	11.46
Work Experience (years)	1	56	15	16	11.11
Education	2	5	4	3.6	0.85

Panel B: Pearson Correlation Matrix

Parameter	*1*	*2*	*3*	*4*	*5*	*6*
1. *Preferred Evaluator*	1					
2. *Fairest Evaluator*	0.63***	1				
3. *Past Discrimination*	0.23*	0.27***	1			
4. *Age (years)*	--0.01	-0.14	0.04	1		
5. *Work Experience*	-0.06	-0.14	-0.07	0.86***	1	
6. *Education*	0.19	0.14	0.00	0.03	-0.06	1

Variable Definitions:

1. *Preferred Evaluator*: participants' responses to the statement, "Which would you rather have evaluating you at ABC Robotics: a human manager or an artificial intelligence algorithm?" Responses were recorded on a five-point scale from "I strongly prefer a human manager to evaluate me" (coded as 1) to "I strongly prefer an artificial intelligence algorithm to evaluate me" (coded as 5).
2. *Fairest Evaluator:* participants' responses to the statement, "Which would you expect to be better at making a fair final evaluation, AI or a human manager?" Responses were recorded on a five-point scale from "AI is much better" (coded as 1) to "A human manager is much better" (coded as 5).
3. *Past Discrimination:* participants' responses to the statement, "I have been subject to discrimination at work." Responses were recorded on a seven-point scale from "Strongly disagree" (coded as 1) to "Strongly agree" (coded as 7).
4. *Age:* participants' self-reported age in years.
5. *Work Experience:* participants' self-reported work experience in years.
6. *Education*: participants' self-reported education level recorded on a five-point scale from "less than a high school degree" (coded as 1) to "higher than a college degree" (coded as 5).

Throughout the paper: *, **, *** denote significance at the, 0.1, 0.05 and <0.01 level.

Table 2: Experiment 2 Descriptive Statistics and Correlation Tables

Panel A: Descriptive Statistics

Parameter	*Min*	*Max*	*Median*	*Mean*	*SD*
Preferred Evaluator	1	5	3	2.91	1.33
Social Intelligence	51	145	89	94.28	17.96
Fairest Evaluator	1	5	3	3.23	1.30
Age (years)	20	69	32	36	10.32
Work Experience (years)	0	47	10	12	9.36
Education	2	5	4	3.73	0.78

Panel B: Correlations

Parameter	*1*	*2*	*3*	*4*	*5*	*6*
1. Preferred Evaluator	1					
2. *Social Intelligence*	-0.28***	1				
3. *Fairest Evaluator*	0.46***	-0.11	1			
4. *Age*	0.04	0.15	-0.10	1		
5. *Work Experience*	-0.12	0.31***	-0.05	0.78***	1	
6. *Education*	0.26***	-0.14	0.22***	-0.04	-0.10	1

Variable Definitions:
2. Social Intelligence: Sum of the responses to 21 items of the Social Intelligence Scale created and validated in Silvera et al. (2001). See appendix for complete scale.

See Table 1, Panel A for all other variable definitions.

Table 3: Main Test of H1

Panel A: *Preferred Evaluator* by *Environment Stability*

	n	*Preferred Evaluator*	SD
Unstable	53	2.43	1.37
Stable	71	2.96	1.21

Panel B: Main Test of H1: ANCOVA

Model: *Preferred Evaluator = Environment Stability + Past Discrimination +* ϵ

	df	*MS*	*p*
Environment Stability	1	11.38	$<0.01^{***}$
Past Discrimination	1	13.97	$<0.01^{***}$
Error	121		

Panel A reports cell sizes, means, and standard deviations of *Preferred Evaluator* for each condition of *Environment Stability,* collected from experiment 1.

Panel B reports the results of an ANCOVA model including *Environment Stability* and *Past Discrimination* as independent variables and *Preferred Evaluator* as the dependent variable. See Table 1 for variable definitions.

Table 4: Mediation Analysis of the Effect of *Environment Stability* on *Preferred Evaluator* While Controlling for *Past Discrimination*

Panel A: Outcome Variable: *AI Context*

	β	SE	*t*	*p*
Environment Stability	0.95	0.35	2.74	<0.01***
Past Discrimination	0.05	0.91	0.54	0.59

Panel B: Outcome Variable: *Human Context*

	β	SE	*t*	*p*
Environment Stability	-0.47	0.30	-1.58	0.12
Past Discrimination	-0.39	0.08	-4.90	<0.01***

Panel C: Outcome Variable: *Preferred Evaluator*

	β	SE	*t*	*p*
Environment Stability	0.24	0.19	1.26	0.21
AI Context	0.22	0.05	4.65	<0.01***
Human Context	-0.36	0.06	-6.48	<0.01***
Past Discrimination	0.03	0.05	0.59	0.55

Panel D: Indirect Effects of *Environment Stability* on *Preferred Evaluator*

	β	SE	*Lower CI*	*Upper CI*
Total	0.38	0.16	0.09	0.70
AI Context	0.21	0.10	0.04	0.44
Human Context	0.17	0.11	-0.04	0.40

We conduct a mediation analysis of the effect of *Environment Stability* on *Preferred Evaluator* through two parallel mediators: *AI Context* and *Human Context*. We include *Past Discrimination* as a covariate. We conduct this analysis using the simultaneous OLS regression method and Model 4 outlined in Hayes (2018). See Figure 7 for a visual depiction of the model and results. Tests are conducted using a 95% confidence interval and 5,000 bootstrap samples. *AI Context* is participants' response to the following PEQ item on a seven-point scale from strongly disagree (coded as 1) to strongly agree (coded as 7): "When casting my vote for who would evaluate me (AI or a human manager), I was concerned that an AI wouldn't be able to fairly consider the circumstances I was in." *Human Context* is participants' response to the following PEQ item on the same scale: "When casting my vote for who would evaluate me (AI or a human manager), I was concerned that a human manager wouldn't be able to fairly consider the circumstances I was in." See Table 1 for other variable definitions.

Table 5: Randomization Checks

Panel A: Experiment 1 *Past Discrimination* between Levels of *Environment Stability*

	within *Stable*	within *Unstable*	Difference	*t*	*p*
Past Discrimination	3.74	3.21	0.53	1.54	0.13

Panel B: Experiment 2 *Past Discrimination* and *Social Intelligence* between Levels of *Workspace*

	within *Shared*	within *Remote*	Difference	*t*	*p*
Past Discrimination	3.91	3.73	0.17	0.55	0.58
Social Intelligence	95.03	93.57	1.46	0.50	0.62

We test for successful randomization of participants by comparing mean levels of measured variables across conditions of our manipulated variables. We find no significant differences in any of the measured variables between conditions, suggesting successful randomization. See Tables 1 and 2 for variable definitions.

Table 6: Main Tests of H2a and H2b

Panel A: The Effect of *Past Discrimination* and *Workspace* on *Preferred Evaluator*

ANCOVA: *Preferred Evaluator = Workspace + Past Discrimination + Past Discrimination*Workspace +* ϵ

	df	*MS*	*F*	*p*
Workspace	1	4.89	3.09	0.08^{*}
Past Discrimination	1	24.95	15.77	$<0.01^{***}$
Past Discrimination Workspace*	1	7.25	4.58	0.03^{**}
Error	151			

Effects of *Past Discrimination* on *Preferred Evaluator* at each level of *Workspace*

	df	*MS*	*F*	*p*
High	1	2.50	1.46	0.23
Low	1	31.53	21.67	$<0.01^{***}$

Panel B: *Preferred Evaluator* by *Workspace* and median split of *Past Discrimination*

	Workspace		
Past Discrimination	*Shared*	*Remote*	Difference
High	3.11	3.59	0.48
Low	2.71	2.47	-0.24
Difference	0.40	1.12	

In Panel A, we conduct an ANCOVA using *Preferred Evaluator* as the dependent variable and *Workspace*, *Past Discrimination,* and the interactive effect as independent variables.

Panel B reports the mean *Preferred Evaluator* of the four groups created by performing a median split of the sample by both *Workspace* and *Past Discrimination*. See Table 1 for variable definitions.

Table 7: Main Tests of H2c and H2d

Panel A: The Effect of *Social Intelligence* and *Workspace* on *Preferred Evaluator*

ANCOVA: *Preferred Evaluator = Social Intelligence + Workspace +Social Intelligence* Workspace + Past Discrimination +* ϵ

	Df	*MS*	*F*	*p*
Social Intelligence	1	6.37	4.07	0.04^{**}
Workspace	1	5.11	3.27	0.07^{*}
Social Intelligence Workspace*	1	5.51	3.53	0.06^{*}
Past Discrimination	1	11.93	7.63	$<0.01^{***}$
Error	150			

Effects of *Social Intelligence* on *Preferred Evaluator* at each level of *Workspace*

	Df	*MS*	*F*	*p*
High SI	1	21.95	15.31	$<0.01^{***}$
Low SI	1	0.97	0.66	0.42

Panel B: *Preferred Evaluator* by *Workspace* and median split of *Social Intelligence*

	Workspace		
Social Intelligence	*Shared*	*Remote*	Difference
High	2.44	2.47	-0.03
Low	3.43	3.34	0.09
Difference	-0.99	-0.87	

In Panel A, we conduct an ANCOVA using *Preferred Evaluator* as the dependent variable and *Workspace*, *Social Intelligence*, and the interactive effect as independent variables. We also include *Past Discrimination* as a covariate.

Panel B reports the mean *Preferred Evaluator* of the four groups created by performing a median split of the sample by both *Workspace* and *Social Intelligence*. See Tables 1 and 2 for variable definitions.

Table 8: Effect of *Environment Stability* and *Past Discrimination* on *Fairest Evaluator*

Panel A: *Fairest Evaluator* by *Environment Stability*

	n	*Fairest Evaluator*	SD
Unstable	53	2.54	1.22
Stable	71	2.92	1.08

Panel B: ANCOVA

Model: *Fairest Evaluator* = *Environment Stability* + *Past Discrimination* + ϵ

	df	*MS*	*F*	*p*
Environment Stability	1	6.38	5.34	0.02^{**}
Past Discrimination	1	14.05	11.76	$<0.01^{***}$
Error	121			

Panel A reports cell sizes, means, and standard deviations of *Fairest Evaluator* for each condition of *Environment Stability*, collected from experiment 1.

Panel B reports the results of an ANCOVA model including *Environment Stability* and *Past Discrimination* as independent variables and *Fairest Evaluator* as the dependent variable. See Table 1 for variable definitions.

Table 9: The Effect of *Past Discrimination* and *Workspace* on *Fairest Evaluator*

Model: *Fairest Evaluator = Workspace + Past Discrimination + Workspace*Past Discrimination +* ϵ

	df	*MS*	*F*	*p*
Workspace	1	0.14	0.08	0.77
Past Discrimination	1	1.59	0.93	0.34
Workspace Past Discrimination*	1	<0.01	<0.01	0.97
Error	151			

We conduct an ANCOVA using *Fairest Evaluator* as the dependent variable and *Workspace*, *Past Discrimination*, and the interactive effect as independent variables. See Table 1 for variable definitions.

Table 10: Effect of *Social Intelligence* and *Workspace* on *Fairest Evaluator*

Panel A: The Effect of *Social Intelligence* and *Workspace* on *Fairest Evaluator*

ANCOVA: *Fairest Evaluator = Social Intelligence* + *Workspace*+ *Social Intelligence*Workspace* + *Past Discrimination* + ϵ

	df	*MS*	*F*	*p*
Social Intelligence	1	2.40	1.44	0.23
Workspace	1	6.80	4.07	0.04^{**}
Social Intelligence Workspace*	1	6.04	3.62	0.06^{*}
Past Discrimination	1	0.18	0.11	0.74
Error	150			

Effects of *Social Intelligence* on *Fairest Evaluator* at each level of *Workspace*

	df	*MS*	*F*	*p*
Shared	1	9.01	5.83	0.02^{**}
Remote	1	0.13	0.07	0.79

Panel B: *Fairest Evaluator* by *Workspace* and median split of *Social Intelligence*

	Workspace		
Social Intelligence	*Shared*	*Remote*	Difference
High	2.49	2.67	-0.18
Low	3.29	2.72	0.57
Difference	-0.80	-0.05	

In Panel A, we conduct an ANCOVA using *Fairest Evaluator* as the dependent variable and *Workspace*, *Social Intelligence*, and the interactive effect as independent variables. We also include *Past Discrimination* as a covariate.

Panel B reports the mean *Fairest Evaluator* of the four groups created by performing a median split of the sample by both *Workspace* and *Social Intelligence*. See Tables 1 and 2 for variable definitions.

Table 11: Moderated Mediation Regression Analysis of the Effect of *Social Intelligence* on *Preferred Evaluator* through *Fairest Evaluator,* Moderated by *Workspace*

Outcome Variable: ***Fairest Evaluator***				
	B	SE	*t*	*p*
Social Intelligence	-0.02	0.01	-2.13	$0.02^{**\dagger}$
Workspace	-2.25	1.12	-2.02	$0.02^{**\dagger}$
*Social Intelligence * Workspace*	0.02	0.01	1.90	$0.03^{**\dagger}$
Past Discrimination	0.02	0.06	0.32	0.74
Outcome Variable: ***Preferred Evaluator***				
	B	SE	*t*	*p*
Social Intelligence	-0.01	0.01	-1.65	$0.05^{**\dagger}$
Fairest Evaluator	0.43	0.07	6.18	$<0.01^{***\dagger}$
Past Discrimination	0.15	0.05	2.88	$<0.01^{***}$
Indirect Effect of ***Social Intelligence*** **on** ***Preferred Evaluator*** **through** ***Fairest Evaluator*** **for each level of** ***Workspace***				
	B	SE	Lower CI	Upper CI
Shared	-0.01	0.004	-0.0160	-0.0007
Remote	<0.01	0.004	-0.0062	0.0091
Index of moderated mediation	Index	SE	Lower	Upper
Workspace	0.0096	0.006	-0.0003	0.0208

We test a moderated mediation model using the simultaneous OLS regression method and Model 7 outlined in Hayes (2018). This approach allows an examination of whether a) there is an indirect effect of the independent variable on the dependent variable through a mediator, and b) whether such an indirect effect is conditional on a moderating variable. In this model, we test whether *Social Intelligence* affects *Preferred Evaluator* through *Fairest Evaluator* and whether this indirect effect is conditional on *Workspace*. Tests are conducted using 5,000 bootstrap samples and a 95% confidence interval. See Tables 1 and 2 for variable definitions.

† One-tailed test consistent with directional prediction.

Appendix: Tromsø Social Intelligence Scale (Silvera, Martinussen, and Dahl 2001)

All responses were recorded on a seven-point Likert scale from “strongly agree” to “strongly disagree.” Items 2, 4, 5, 8, 11, 12, 13, 15, 16, 20, 21 are reverse coded.

1. I can predict other peoples’ behavior.
2. I often feel that it is difficult to understand others’ choices.
3. I know how my actions will make others feel.
4. I often feel uncertain around new people who I don’t know.
5. People often surprise me with the things they do.
6. I understand other peoples’ feelings.
7. I fit in easily in social situations.
8. Other people become angry with me without me being able to explain why.
9. I understand others’ wishes.
10. I am good at entering new situations and meeting people for the first time.
11. It seems as though people are often angry or irritated with me when I say what I think.
12. I have a hard time getting along with other people.
13. I find people unpredictable.
14. I can often understand what others are trying to accomplish without the need for them to say anything.
15. It takes a long time for me to get to know others well.
16. I have often hurt others without realizing it.
17. I can predict how others will react to my behavior.
18. I am good at getting on good terms with new people.
19. I can often understand what others really mean through their expression, body language, etc.
20. I frequently have problems finding good conversation topics.
21. I am often surprised by others’ reactions to what I do.