Does AI Supervision "Benefit" the Organization? Experimental Evidence based on Employees' Extra-Role Behaviors

Yunshil Cha
University of New Hampshire
10 Garrison Avenue, Durham, NH 03824
yunshil.cha@unh.edu

Cody Lu
University of Massachusetts Lowell
72 University Avenue, Lowell, MA 01854
Cody_Lu@uml.edu

Corresponding author: Yunshil Cha, University of New Hampshire, Peter T. Paul College of Business, 10 Garrison Avenue, Durham, NH, USA. Phone: (603) 862-0612.

Funding: We gratefully acknowledge the financial support provided by the Accounting Department at the University of Massachusetts Lowell.
Data Availability: available upon request.
Keywords: artificial intelligence; AI supervisors; extra-role behaviors; performance evaluation systems; quantification; humanizing technology

Does AI Supervision "Benefit" the Organization?
Experimental Evidence based on Employees' Extra-Role Behaviors

Abstract: The utilization of "AI supervisors" in organizations to manage and evaluate employee performance has risen in recent years. In this study, we employ two experiments to examine how AI supervisors influence subordinates' willingness to engage in extra-role behaviors. In the first experiment, we predict and find that under a purely objective performance evaluation system, supervisor type does not significantly affect subordinates' willingness to engage in extra-role behaviors, but that under a mixed performance evaluation system that contains subjective criteria, AI supervisors reduce subordinates' willingness to engage in extra-role behaviors. This relationship is driven by subordinates' beliefs that AI supervisors will quantify their extra-role contributions and reciprocate less in response to their extra-role behaviors. In our second experiment, we predict and find that humanizing AI supervisors (i.e., by giving them a name) can mitigate the negative impact of AI supervisors on subordinates' willingness to engage in extra-role behaviors.

Keywords: artificial intelligence; AI supervisors; extra-role behaviors; performance evaluation systems; quantification; humanizing technology

1. Introduction

Technological innovations and the availability of big data continue to fundamentally change the role of artificial intelligence (AI) in organizations, as well as the nature of interactions between humans and AI in the workplace (Singla et al., 2024). One such development in recent years is the emergence of "AI supervisors," who have the authority to exercise influence over human employees (Bol et al., 2025; Brown et al., 2026; Cha, 2025; Lanz et al., 2024; Wesche and Sonderegger, 2019).[1] Similar to human supervisors, AI supervisors provide instruction to subordinates, evaluate performance, and make promotion and retention decisions (Brown et al., 2026; Höddinghaus et al., 2021; Lecher, 2019; Möhlmann et al., 2021; Parent-Rocheleau and Parker, 2022; Resume Builder, 2025). A survey of 1,342 U.S. managers reports that approximately 60 percent of respondents rely on AI to perform these supervisory functions, and 20 percent delegate to AI the authority to make final decisions without human input (Resume Builder, 2025). Beyond survey evidence, AI in supervisory roles is evident in several prominent organizations. Uber's algorithm directs drivers on when and where to work, monitors drivers' behavior through real-time GPS and acceptance-rate data, and unilaterally deactivates drivers' accounts for non-compliance, all without human intervention (Möhlmann et al., 2021). At Amazon, the Associate Development and Performance Tracker (ADAPT) monitors every worker's productivity and automatically generates warnings and even terminations without any input from human supervisors, with the system having fired roughly 300 workers at a single facility within a thirteen-month period (Lecher, 2019). UPS's On-Road Integrated Optimization

[1]Scholars have used different terms to refer to AI's exertion of influence over human employees, including algorithmic management (Harms and Han, 2019; Lee, 2018; Newman et al., 2020), automated leadership (Höddinghaus et al., 2021), and AI supervisors (Cui and Yasseri, 2026; Lanz et al., 2024). We view these terms interchangeably, but following more recent studies, we employ the term "AI supervisor" throughout the remainder of our study.

and Navigation (ORION) system dictates where drivers go by providing them with an optimized route (Holland et al., 2017). Moreover, following ORION's implementation, the company has changed how it evaluates drivers. For instance, rather than evaluating drivers based on their average delivery time, UPS now evaluates employees by measuring how closely drivers follow the algorithm's optimized route (Tabansi, 2025). In both these anecdotal examples and the survey evidence, the AI algorithms are not simply assisting or supplementing human managers; rather, the algorithms *are* the managers.

Prior research examining the impact of algorithms in a decision support role documents differences in how individuals perceive algorithmic versus human decisions across dimensions such as objectivity (Logg et al., 2019), competence and credibility (Chen et al., 2022), fairness and trustworthiness (Lee, 2018), task complexity (Hodge et al., 2021), and uniqueness (Longoni et al., 2019). The shift in the use of algorithms from a decision support role to a supervisory role may change how employees perceive algorithms and in turn, how they respond to performance evaluations conducted by AI supervisors. This backdrop leads to our primary research question: how does AI supervision affect subordinates' willingness to engage in extra-role behaviors, and is this relationship influenced by the design of firms' performance evaluation system?

Extra-role behaviors (or organizational citizenship behavior) refer to employees' discretionary contributions to the organization that go above and beyond their core job duties (Becton et al., 2008; Organ, 1988), and is a topic that has received attention in the accounting literature (e.g., Bol et al., 2026; Burney et al., 2009; Cheng and Coyte, 2014). Although extra-role behaviors such as helping coworkers and sharing knowledge are typically not explicitly recognized by firms' formal reward systems, these activities improve firms' productivity and efficiency, which help reduce costs (Organ, 1988; Podsakoff et al., 1997; Podsakaoff et al.,

2009). Consequently, firms have strong incentives to motivate such behavior from employees (Becton et al., 2008; Bolino et al., 2010; He et al., 2021; Podsakoff et al., 2009).

Prior research has documented mixed evidence on the pros and cons of algorithmic decision-making in performance evaluation. While some studies highlight the benefits of such decisions, including increased perceptions of unbiasedness and objectivity (Brown et al., 2026; Cha, 2025; Pethig and Kroenung, 2023), other studies show that individuals perceive algorithmic decisions as reductionistic and decontextualized (Bol et al., 2025; Newman et al., 2020). Specifically, Newman et al. (2020) argue and find that employees associate algorithms with a tendency to reduce the qualitative aspects of their performance into quantifiable metrics, a phenomenon termed "quantification." As a result, performance evaluations made by algorithms are perceived as more incomplete and reductionistic relative to those made by humans. In contrast, human managers value employees who engage in extra-role behaviors and reflect such activities in their performance ratings and reward allocation decisions (Lefkowitz, 2000; Podsakaoff et al., 1993; Podsakaoff et al., 2009; Werner, 1994). Taken together, these findings suggest that relative to algorithmic managers, human managers are expected to be more likely to incorporate qualitative aspects of performance information, such as employees' discretionary contributions beyond their core job responsibilities, into performance evaluations.

To understand how differential perceptions of AI versus human supervisors influence subordinates' willingness to engage in extra-role behaviors, we draw on social exchange theory, which posits that individuals act to maximize their net benefit through a psychological cost-benefit analysis (Blau, 1964; Cook et al., 2013; Cropanzano and Mitchell, 2005; Thibaut and Kelley, 1959). In other words, the theory predicts that subordinates expect recognition for costly actions (e.g., extra-role behaviors). Moreover, when the amount of recognition for their costly

actions is perceived to be inadequate, subordinates will in turn become less willing to engage in such actions (Bordia et al., 2008; Restubog et al., 2008). Consistent with this prediction, Cheng and Coyte (2014) show that subordinates perceive a higher overall net benefit of engaging in extra-role behaviors when evaluated under a performance evaluation system that allows for subjective/flexible weighting of performance measures than under a performance evaluation system that is strictly formula-based with pre-defined weights, so long as the firm communicates how subordinates' inputs can contribute to strategic goals. This finding suggests that, since allowing for subjectivity provides supervisors with the discretion and opportunity to reward subordinates for demonstrating extra-role behaviors, subordinates are in turn more willing to pursue extra-role behaviors.

Building on this framework, we conjecture that regardless of supervisor type, subordinates are less willing to engage in extra-role behaviors when evaluated based solely on pre-defined objective criteria, given that these types of performance evaluation systems leave little room for supervisors to recognize extra-role contributions. In contrast, when subordinates are evaluated under a "mixed" performance evaluation system that contains subjective criteria, we expect that their willingness to engage in extra-role behaviors will depend on supervisor type. We posit that because subordinates will expect AI supervisors to overly quantify the qualitative aspects of their performance (Newman et al., 2020), such as the incremental effort required to engage in extra-role behaviors, subordinates will be less likely to believe that their extra-role behaviors will be fully recognized and/or reciprocated by AI supervisors. In turn, subordinates will be less willing to engage in extra-role behaviors. Thus, we predict that subordinates under AI supervision will be less willing to engage in extra-role behaviors than subordinates under human supervision when the performance evaluation system contains subjective criteria.

Given the importance of extra-role behaviors, our initial prediction naturally begs the question of how organizations can encourage greater extra-role behaviors from subordinates under AI supervision. As discussed above, subordinates' lower willingness to engage in extra-role behaviors stems from their perception that the AI supervisor will overly quantify their qualitative contributions and thus, not fully recognize them in the way that a human supervisor would. We argue that one way to encourage more extra-role behaviors from subordinates is to humanize (or anthropomorphize) AI supervisors. Although there are different methods with which to humanize technology, both prior studies and anecdotal evidence have shown that one of the most practical ways to do so is to give the technology a humanlike name (El Halabi and Trendel, 2024; Hodge et al., 2021; Johnson, 2017; Zhou and Wang, 2025), such as "Watson" (IBM), "Einstein" (Salesforce), and "Joule" (SAP). Research on human-computer interactions finds that people typically associate more positive feelings, such as friendliness and engagement, with technologies that they find to be more humanlike (e.g., El Halabi and Trendel, 2024; Hodge et al., 2021; Qiao and Song, 2025; Waytz et al., 2014). These findings insinuate that naming AI supervisors should cause subordinates to perceive AI supervisors as more humanlike. Based on our theory, this will in turn strengthen subordinates' belief that their AI supervisor will recognize and reward them for extra-role behaviors when evaluated under mixed evaluation systems that include subjective criteria. This argument leads us to predict that humanizing AI supervisors, by giving them a name, will attenuate the negative effect of AI supervision on subordinates' willingness to engage in extra-role behaviors.

To test our predictions, we conduct two online experiments using Prolific participants acting in the role of a management consultant. Our first experiment is a 2 × 2 between-participants design that examines the impact of supervisor type and performance evaluation

system on extra-role behaviors. We manipulate supervisor type by informing participants that their supervisor is either a person or AI system. Following Kunz (2015), we manipulate performance evaluation system by informing participants that they are evaluated based either exclusively on objectively measurable criteria (i.e., an objective performance evaluation system) or on both objectively measurable criteria and subjective criteria (i.e., a mixed performance evaluation system). Our second experiment is a 2 × 2 between-participants design that examines the impact of supervisor type and the level of humanization on extra-role behaviors under a mixed performance evaluation system. We manipulate supervisor type in the same manner as in Experiment 1. We manipulate the level of humanization by either providing or not providing a name for the supervisor. In both experiments, participants assess their willingness to engage in extra-role behaviors after being presented with their hypothetical scenario (which contains our manipulated information).

Consistent with our predictions, we find that supervisor type does not influence subordinates' willingness to partake in extra-role behaviors under a purely objective performance evaluation system. However, we find that subordinates under AI supervision are less willing to engage in extra-role behaviors than those under human supervision when the performance evaluation system contains subjective criteria. Additional mediation analysis suggests that this effect is driven by subordinates' perception that AI supervisors will reduce their extra-role contributions into quantifiable metrics, leaving less room for recognition and reciprocation in performance evaluations. We also find that naming AI supervisors helps to mitigate the negative effect of AI supervision on subordinates' willingness to engage in extra-role behaviors, which suggests that humanizing AI supervisors can enable organizations to utilize AI supervisors while limiting the associated deleterious impact on subordinates' extra-role behaviors.

Our study provides several contributions. First, we extend an emerging stream of literature on algorithmic performance evaluations (Bol et al., 2025; Newman et al., 2020). Newman et al. (2020) show that employees perceive algorithmic evaluations as reductionistic for two reasons: algorithms tend to compress qualitative performance into numerical metrics (quantification) and fail to consider qualitative performance within a broader context (decontextualization). Building on the notion that AI is less able to contextualize performance information, Bol et al. (2025) show that employees prefer human evaluators over AI evaluators in unstable environments, where contextualization is more important. We focus on the other source of reductionism, quantification, and examine whether employees reduce their willingness to engage in extra-role behaviors when evaluated by AI supervisors. We find that under mixed performance evaluation systems that contain subjective criteria, employees evaluated by AI supervisors are in fact less willing to engage in extra-role behaviors than those evaluated by human supervisors, because they believe that AI supervisors are more likely to quantify their extra-role contributions and therefore, to recognize and reciprocate such contributions less during performance evaluations. Our study offers insights to organizations seeking to understand how AI supervisors can be effectively integrated into workplace performance evaluation systems.

Second, our findings extend accounting research on extra-role behaviors, which remains relatively underexplored despite its well-documented benefits to organizations (Podsakoff et al., 1997; Podsakoff et al., 2009). While extra-role behaviors are traditionally viewed as activities that are not explicitly or formally recognized by reward systems (Organ, 1988), prior studies in accounting have acknowledged and documented that accounting systems can be designed in a way that implicitly or indirectly induces extra-role behaviors (e.g., Bol et al., 2026; Burney et al., 2009; Cheng and Coyte, 2014). Most relevant to our study is the finding that introducing

subjectivity into performance evaluation systems can encourage extra-role behaviors (Cheng and Coyte, 2014). Our findings extend this notion further by demonstrating that subjectivity may not effectively incentivize extra-role behaviors if employees are evaluated by AI supervisors. Given the importance of extra-role behaviors on organizational success, our study provides insights to organizations on how to weigh both performance evaluation system design and supervisor type when it comes to motivating extra-role behaviors.

Finally, we contribute to the broad literature on human-algorithm interactions. Although humanizing technology has generally been advocated for in prior research (e.g., Chaminade et al., 2007; El Halabi and Trendel, 2024; Gong, 2008; Qiao and Song, 2025; Waytz, et al., 2014), contrasting sentiments have argued that there may also be disadvantages to humanization (Hodge et al., 2021; Nan et al., 2006). Most notably in accounting research is the finding that humanizing robo-advisors can cause investors to perceive the robo-advisors as less competent, particularly when a task is more complex, which in turn reduces investors' reliance on the investment recommendations provided by robo-advisors (Hodge et al., 2021). Our findings complement these studies by illustrating the value of humanizing technology when the technology acts as the direct supervisor of a human employee. Specifically, we show that employees may perceive more humanlike AI supervisors as more likely to recognize and reward them for extra-role behaviors, which in turn can increase employees' willingness to engage in extra-role behaviors.

2. Background and Hypothesis Development

2.1. Extra-Role Behavior

Extra-role behavior is defined as "individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and that in the aggregate promotes the efficient and effective functioning of the organization" (Organ, 1988, p. 4). Examples of such

behaviors include helping others with heavy workloads, sharing knowledge, and taking time to listen to co-workers' problems (Cheng and Coyte, 2014; Williams and Anderson, 1991). Extra-role behaviors enhance organizational functioning by improving coordination, strengthening team activities and cohesion, and fostering innovation (Marinova et al., 2018; Podsakoff et al., 2000). Moreover, prior research has also documented that employees who engage in extra-role behaviors derive benefits in other areas such as receiving more favorable evaluations from their supervisors on performance dimensions related to their primary job functions and higher likelihoods of a promotion (Lefkowitz, 2000; Marinova et al., 2013; Podsakoff et al., 1993; Podsakoff et al., 2009). Hence, extra-role behaviors are beneficial to both organizations and employees, giving organizations strong incentives to encourage such behaviors and give employees reasons to engage in them (Bolino et al., 2010; He et al., 2021).

2.2. Human versus AI Supervisors

AI technologies have been rapidly adopted by organizations to support managerial decision-making. One area where AI use is becoming particularly prominent is in conducting performance evaluations. Algorithms were initially adopted as decision-aid tools to assist human managers in performance evaluations, with human managers retaining final decision-making authority. For example, Lin et al. (2026) find that managers are less willing to use a performance evaluation algorithm when the algorithm advises a low rating for subordinates. However, when managers have discretion to adjust the algorithm's computation process, managers are more willing to use the algorithm even when the algorithm advises a low rating.

While the use of algorithms as decision-aid tools remains prevalent, algorithms are increasingly assuming autonomous supervisory roles in some organizations such as Amazon and Uber. Firms originally deployed AI supervisors at the middle management level to bridge upper

management and lower-level employees (Wesche and Sonderegger, 2019), but AI supervisors have more recently been given greater control over performance evaluations without human intervention (Lecher, 2019; Lee et al., 2015; Möhlmann et al., 2021). Reflecting this development, a growing body of research has begun to examine how employees perceive and respond to algorithmic evaluators, yielding mixed evidence that documents both algorithm appreciation and algorithm aversion.

Studies documenting algorithm appreciation suggest that individuals tend to view algorithmic evaluators favorably relative to human evaluators, particularly in contexts where bias is a concern. Pethig and Kroenung (2023) show that female employees prefer an algorithmic evaluator when the alternative is a male evaluator. Similarly, Cha (2025) finds that employees perceive an algorithmic evaluator as more trustworthy than a human evaluator when they are vulnerable to evaluation bias. Brown et al. (2026) show that employees from underrepresented minority backgrounds view White male evaluators less favorably than AI evaluators. Taken together, these studies suggest that individuals tend to perceive algorithmic evaluators as less biased than their human counterparts.

Despite these findings of algorithm appreciation, a contrasting body of research documents algorithm aversion among employees. Lee (2018) finds that subordinates generally perceive automated supervision to be less fair and trustworthy, particularly on tasks that are more human-oriented. Relatedly, algorithms are viewed as less adaptable and benevolent than human supervisors because they lack the ability to feel emotion and empathy (Bigman and Gray, 2018; Höddinghaus et al., 2021). Newman et al. (2020) find that employees perceive algorithmic evaluators as less fair, viewing algorithmic evaluations as fundamentally reductionistic through quantification and decontextualization. On the other hand, employees perceive human evaluators

as more likely to apply qualitative and holistic assessments of their performance. Bol et al. (2025) further find that employees prefer a human evaluator over an AI evaluator in stable operating environments where contextual information is readily available. Taken together, these studies suggest that employees may also view algorithmic evaluators less favorably than human evaluators, particularly when tasks involve and require human qualities.

2.3. Performance Evaluation Systems

Objective performance measures, such as sales or output, capture only contractible information, which therefore offer an incomplete view of an employee's overall performance (Datar et al., 2001; Feltham and Xie, 1994). To address this limitation, managers typically incorporate subjectivity into their evaluations by combining subjective and objective performance measures (Bol, 2008). Allowing subjectivity gives managers the discretion to consider noncontractible contributions, i.e., those that matter to organizational performance but cannot be specified in formal contracts or captured by pre-defined objective criteria, such as effort and extra-role behaviors. In our study, we focus primarily on the differences between a performance evaluation system that contains only objective measures and a mixed system that contains both objective and subjective measures (Kunz, 2015).

Despite the notion that extra-role behaviors correspond to actions that are not directly or explicitly recognized (Organ, 1988; Organ and Ryan, 1995), prior studies have shown that performance evaluation systems can be configured in a way that indirectly or implicitly induces extra-role behaviors from employees. For example, Burney et al. (2009) provide survey evidence documenting that incorporating strategic causal models into performance management systems can increase willingness to engage in extra-role behaviors. Drake et al. (1999) and Taylor (2006) find that on tasks with a high degree of interdependence, utilizing group-based incentive schemes

can promote knowledge sharing and helping behaviors (two forms of extra-role behaviors) amongst team members. Most relevant to our study, Cheng and Coyte (2014) find that an incentive scheme that allows subjective weighting, when complemented with a strategy map that communicates how employees' knowledge, skills, and values contribute to the firm's strategic goals, can increase employees' tendency and willingness to pursue discretionary extra-role behavior, relative to a purely formula-based incentive scheme. Thus, these findings suggest that employees' willingness to engage in extra-role behaviors may be influenced by the design of performance evaluation systems. Our study builds on the idea that subjectivity in performance evaluation systems can in theory motivate extra-role behaviors to investigate the interaction between supervisor type and performance evaluation system design on extra-role behaviors, which we expound on next.

2.4. Hypothesis Development

Prior research suggests that algorithms are perceived to be at least as capable as humans at performing quantitative tasks in objective domains. Logg et al. (2019) find that individuals rely on algorithms more than humans for numerical tasks with an objectively correct answer. Similarly, Castelo et al. (2019) show that people trust algorithms for tasks that seem objective, namely those involving facts that are quantifiable and measurable. Lee (2018) further finds that people perceive algorithms to perform mechanical tasks as well as human managers. These findings suggest that subordinates believe that human and AI supervisors do not differ significantly in how they evaluate performance when the evaluation is based on pre-defined objective criteria. As purely objective evaluation criteria provide little room for evaluator discretion, supervisor type is unlikely to influence subordinates' perceptions of how their performance will be evaluated.

Social exchange theory argues that social behaviors in the context of relationships are determined based on a psychological process of cost-benefit analysis, whereby individuals typically act in a manner that maximizes their overall return from engaging in a particular behavior (Blau, 1964; Cook et al., 2013; Cropanzano and Mitchell, 2005; Thibaut and Kelley, 1959). In workplace settings, social exchange theory posits that subordinates will expect that costly actions (e.g., putting forth effort, partaking in extra-role behaviors) will be recognized or reciprocated by their supervisor. For example, employees engage in extra-role behaviors to enhance their image (Bolino, 1999), and employees labeled as good citizens receive greater rewards from their supervisors (Eastman, 1994). When reciprocation is perceived to be lacking, subordinates will experience dissatisfaction with the exchange relationship, which can lead to a lower willingness to engage in costly actions (Bordia et al., 2008; Cheng and Coyte, 2014; Restubog et al., 2008). Under a purely objective evaluation system that is strictly based on pre-defined measures, neither human nor AI supervisors have discretion to differentially recognize or reciprocate subordinates' extra-role contributions. Thus, we conjecture that supervisor type will not significantly affect subordinates' willingness to engage in extra-role behaviors under purely objective evaluation systems.

Under a mixed evaluation system that contains subjective criteria, however, we expect supervisor type to significantly affect subordinates' willingness to engage in extra-role behaviors. Prior research suggests that individuals perceive algorithmic evaluators as less capable than human evaluators of assessing performance dimensions that require subjective judgment. Lee (2018) finds that algorithmic decisions are perceived as less fair and trustworthy than human decisions on tasks that require human skills, as algorithms are viewed as lacking intuition and subjective judgment capabilities. Newman et al. (2020) argue that algorithmic reductionism

manifests in two ways. First, algorithms excessively quantify qualitative dimensions of performance into numerical metrics, a process through which important performance information can be lost. Second, algorithms decontextualize performance information, evaluating employee performance in isolation rather than within the broader organizational context in which they occur. Together, these two forms of reductionism lead employees to view evaluations conducted by AI supervisors as based on incomplete information and therefore less fair than evaluations conducted by human supervisors. In contrast, human supervisors are capable of holistically assessing employee performance by taking both quantitative and qualitative information into account. Thus, under a mixed evaluation system that contains both objective and subjective evaluation criteria, subordinates will expect a human supervisor to evaluate both criteria holistically, whereas subordinates will expect an AI supervisor to quantify qualitative dimensions that fall under the subjective criteria, which may cause subordinates' extra-role contributions to be overlooked in the evaluation process. Consequently, subordinates evaluated by an AI supervisor (rather than a human supervisor) are less likely to believe that their extra-role contributions will be recognized under a mixed performance evaluation system. As a result, we predict that subordinates evaluated by an AI supervisor are less willing to engage in extra-role behaviors than those evaluated by a human supervisor. The pattern of our prediction is presented in Figure 1, Panel A. Based on this discussion, we pose the following hypothesis:

H1: Subordinates' willingness to engage in extra-role behaviors is lower under AI supervision than under human supervision when the performance evaluation system contains subjective criteria, but supervisor type does not influence subordinates' willingness to engage in extra-role behaviors when the performance evaluation system contains purely objective criteria.

[Insert Figure 1 about here]

Given our predicted negative relationship between AI supervisors and subordinates' willingness to engage in extra-role behaviors, it is important to identify conditions in which such

adverse effects can be mitigated. We conjecture that one such condition involves humanizing AI supervisors. The process of humanizing technology (also referred to as anthropomorphism) involves making technological artifacts more humanlike. While there are different ways to humanize a technological artifact, one common approach is to give the technology a name (El Halabi and Trendel, 2024; Hodge et al., 2021; Johnson, 2017; Zhou and Wang, 2025).[2] One reason why companies may name their technologies is because people perceive humanlike technology to be more friendly and engaging, which in turn boosts their willingness to purchase and utilize the technology (El Halabi and Trendel, 2024; Hodge et al., 2021; Johnson, 2017). Prior studies on human-computer interactions echo this sentiment and find that people typically associate more positive feelings with technologies that they find to be more humanlike (e.g., Chaminade et al., 2007; El Halabi and Trendel, 2024; Gong, 2008; Qiao and Song, 2025; Waytz et al., 2014). For instance, robot anthropomorphism has been shown to increase customer satisfaction and repurchase intent of the robot-provided service (El Halabi and Trendel, 2024; Qiao and Song, 2025). In an internal audit setting, Pickard et al. (2020) find that humanizing virtual embodied conversational agents by giving them a voice and humanlike face can improve disclosure quality and increase employees' willingness to disclose internal control violations.[3]

The implications of these prior studies lead us to believe that naming AI supervisors should similarly cause subordinates to perceive their supervisors as more humanlike. Castelo et al. (2019) find that increasing the human-likeness of algorithms increases individuals' reliance on the algorithms for subjective tasks. This suggests that humanizing AI supervisors leads

[2] Other examples of ways to humanize technology include giving it a voice or an avatar (Ganbold et al., 2022; Pickard et al., 2020).

[3] A separate stream of research in accounting and finance focuses on how humanizing robo-advisors influences investors' reliance on the advice and their investment behavior (e.g., Baek and Kim, 2023; Ganbold et al., 2022; Hodge et al., 2021). Although these studies are also interested in documenting the consequences of anthropomorphism, the theoretical framework of these studies typically centers around the tradeoff between humanization and the perceived competence of robo-advisors, which in turn influences investors' behavior.

individuals to perceive them as more capable of exercising subjective judgment in performance evaluations. Based on our theory, this then will cause subordinates to strengthen their belief that the AI supervisor will be less likely to quantify, and more likely to recognize and reward them for, extra-role behaviors (under a mixed performance evaluation system that contains subjective criteria). Hence, we argue that when evaluated based on a mixed performance evaluation system, subordinates are more willing to engage in extra-role behaviors when their AI supervisor has a name than when their AI supervisor does not have a name. The pattern of our prediction is presented in Figure 1, Panel B, and our formal hypothesis is summarized as follows:[4]

H2: Humanizing an AI supervisor by giving it a name will *increase* subordinates' willingness to engage in extra-role behaviors under a mixed performance evaluation system that contains subjective criteria.

3. Experiment 1: Method and Results

Our first experiment utilizes a 2 × 2 between-participants design.[5] We manipulate supervisor type by informing participants that their supervisor is either a human or an AI system. We manipulate performance evaluation system by informing participants that their performance is either evaluated based solely on predefined, objectively measurable criteria (i.e., an objective evaluation system) or based on both objective and subjective criteria (i.e., a mixed evaluation system). We further expand on our experimental methodology in the following subsections.

3.1. Participants

We recruited 200 participants from Prolific for our study. We required participants to be fluent in English, located in the United States, and to be full-time employees. On average,

[4] Although Figure 1, Panel B depicts the full interaction pattern, H2 is stated and tested as a simple effect. We retain the 2 × 2 design to replicate the Experiment 1 prediction that under a mixed performance evaluation system, participants evaluated by an unnamed AI supervisor are less willing to engage in extra-role behaviors than those evaluated by an unnamed human supervisor. Our primary interest, however, is the simple effect of whether humanizing the AI supervisor increases participants' willingness to engage in extra-role behaviors.

[5] This experiment received Institutional Review Board approval at the authors' institutions.

participants were 41 years old and had 20 years of work experience. To ensure the quality of our sample, we also required participants to have an approval rate of at least 99% and to have made a minimum of 20 submissions on Prolific. We also find that 49% of our sample were male and that 95% of our sample indicated they receive some form of performance evaluation at work. Hence, these demographic characteristics suggest that the participants in our sample are suitable for testing our research questions. Participants were paid $2 for completing our experiment and took an average of 7.08 minutes, resulting in an average hourly wage of $16.94.

3.2. Procedure and Manipulations

After accepting the task on Prolific, participants were redirected to Qualtrics, where they provided their consent to participate in our study and were randomly assigned to one of four experimental conditions. We instructed participants to assume the role of a management consultant at Rivion Consulting, a hypothetical consulting firm. After viewing a description and diagram of the organizational hierarchy at Rivion Consulting, participants learned of their supervisor to whom they report. In all conditions, participants were told that their supervisor is responsible for evaluating their performance and providing feedback to them. To manipulate *Supervisor Type*, we tell participants in the *Human* (*AI*) *Supervisor* condition that their supervisor is a person (AI system). The organizational hierarchy diagram also reflected this manipulation and reinforced to participants that their supervisor is either a person or AI system. We provide a depiction of the *Supervisor Type* manipulation in Appendix A.

Participants are then informed about how their performance is evaluated and how their compensation is determined. To manipulate *Performance Evaluation System*, we adapt the approaches used in Kunz (2015). In the *Objective Evaluation System* condition, participants were told that their performance is evaluated *solely* based on the following four predefined

performance criteria: (1) billable hours, (2) average billing rate charged to clients, (3) total revenue from new project work requested by clients, and (4) client evaluation ratings. We then reiterated to participants that they would be evaluated exclusively based on these objectively measurable criteria. In the *Mixed Evaluation System* condition, participants were told that in addition to the same four predefined criteria, their performance is also evaluated based on subjective criteria, such as leadership skills, contributions to teamwork, and potential for growth. In all conditions, participants were informed that their salary and bonus depend to a significant degree on the outcome of their performance evaluation. We provide a depiction of the *Performance Evaluation System* manipulation in Appendix B.

After being exposed to the manipulations, participants answered a series of questions pertaining to their willingness to engage in extra-role behaviors, expected reciprocity from their supervisor, perceptions of the quantification and decontextualization of the performance evaluation process, manipulation checks, and relevant demographics.

3.3. Dependent Variable

Following prior studies (Burney et al., 2009; Cheng and Coyte, 2014; Williams and Anderson, 1991), our dependent variable is participants' *Willingness to Engage in Extra-Role Behaviors*, operationalized using an established measure of organizational citizenship behavior (OCB). Specifically, we asked participants to rate their willingness to devote time to each of five tasks adapted from Williams and Anderson's (1991) OCB-I measure on a seven-point scale (1 = *Strongly disagree*, 7 = *Strongly agree*).[6] These five tasks capture participants' willingness to help

[6] The OCB-I measure captures employees' performance of extra-role behaviors that "have a specific individual as the target" (Williams and Anderson, 1991). Williams and Anderson (1991) also developed the OCB-O measure. In our study, we also solicited participants' willingness to take part in seven activities adapted from the OCB-O measure. The seven activities corresponded to participants' willingness to maintain above-the-norm attendance at work, give advance notice when unable to come to work, refrain from taking undeserved work breaks, limit the amount of time spent on personal phone conversations, refrain from complaining about insignificant things at work, conserve and protect organizational property, and adhere to informal rules devised to maintain order. Although we

co-workers who have been absent, to help co-workers who have heavy workloads, to listen to co-workers' problems and worries, to take a personal interest in other employees, and to pass along information to co-workers. We conveyed to participants that engaging in these tasks may be costly by reminding them that these tasks are not part of their assigned responsibilities at Rivion Consulting and informing them that devoting time to these tasks during regular work hours may delay client projects, which would require them to spend extra time outside of their regular work hours to avoid such delays. Participants indicated their willingness to engage in these tasks on a seven-point scale. We find that the Cronbach's alpha for these five questions is 0.91, which exceeds the recommended threshold of 0.70 (Nunnally, 1978), indicating that the five items are reliable and adequately capture a single underlying construct. We therefore aggregate participants' responses into a single measure by averaging their responses across the five items, with higher values indicating greater willingness to engage in extra-role behaviors.

3.4. Results

3.4.1. Manipulation Checks

To assess the effectiveness of our *Supervisor Type* manipulation, we asked participants to identify which of the following was true in their scenario: "My supervisor was a person" or "My supervisor was an AI system" (Hodge et al., 2021). All our participants correctly answered this question, indicating a successful manipulation of *Supervisor Type*. To assess the effectiveness of our *Performance Evaluation System* manipulation, participants rated their agreement with the following statement in their scenario (on a seven-point scale): "Your performance is *only* measured on the basis of the predefined objectively measurable performance criteria" (Kunz,

document directionally similar results using the OCB-O measure, our results are not statistically significant at conventional levels (i.e., $p > 0.1$). Prior research notes that OCB-O differs from OCB-I in that it reflects generalized compliance, capturing conscientious adherence to organizational norms and expectations rather than discretionary contributions (Williams and Anderson, 1991).

2015). Consistent with our expectations, participants in the *Objective Evaluation System* condition agreed more strongly with this statement than participants in the *Mixed Evaluation System* condition (6.69 versus 1.40, $p < 0.01$), indicating a successful manipulation of *Performance Evaluation System*.[7]

Given the importance to our study of participants understanding how their performance is evaluated, we also administered an attention check question immediately after our manipulations to assess participants' comprehension of the provided information. Specifically, participants were asked whether their supervisor (or the AI system) evaluates their performance *solely* based on the predefined objectively measurable criteria (Yes or No). We find that fourteen participants failed to correctly answer this question, indicating that they did not carefully attend to or adequately comprehend our experimental materials. Therefore, we exclude these participants from subsequent analyses, yielding a final sample of 186 participants.[8]

3.4.2. Test of H1

H1 predicts that subordinates will be less willing to engage in extra-role behaviors when they are evaluated by an AI supervisor than a human supervisor, and that this difference will be more pronounced when the performance evaluation system contains both objective and subjective criteria than when the performance evaluation system relies solely on objective criteria. Table 1, Panel A presents cell sizes, means, and standard deviations for *Willingness to Engage in Extra-Role Behaviors* across the four experimental conditions, and Figure 2 graphically depicts the pattern of cell means by condition. The pattern of cell means observed in Figure 2 is visually consistent with our predicted pattern in Figure 1, Panel A. Table 1, Panel B

[7] All reported p-values are two-tailed unless otherwise specified.
[8] We find that the results and inferences of our subsequent analyses are unchanged when we retain all participants in our sample.

presents an analysis of variance (ANOVA) with *Willingness to Engage in Extra-Role Behaviors* as the dependent variable. Consistent with H1, we find a significant *Supervisor Type* × *Performance Evaluation System* interaction effect ($F_{1,182} = 3.75$, $p = 0.05$). Results also reveal a significant main effect of *Performance Evaluation System* ($F_{1,182} = 13.07$, $p < 0.01$), indicating that participants in the *Mixed Evaluation System* condition are more willing to undertake extra-role behaviors than participants in the *Objective Evaluation System* condition, and a marginally significant main effect of *Supervisor Type* ($F_{1,182} = 3.21$, $p = 0.07$), indicating that participants with AI supervisors are less willing to undertake extra-role behaviors than participants with human supervisors. Both main effects are consistent with our expectations. Table 1, Panel C presents follow-up tests of simple effects to help better understand the nature of our documented interaction. In the *Mixed Evaluation System* condition, participants with an AI supervisor are less willing to engage in extra-role behaviors than participants with a human supervisor (4.78 versus 5.52, $p = 0.01$). Conversely, in the *Objective Evaluation System* condition, *Supervisor Type* does not affect participants' willingness to engage in extra-role behaviors (4.45 versus 4.42, $p = 0.92$). Overall, our documented findings provide evidence consistent with H1.

[Insert Table 1 and Figure 2 about here]

3.4.3. Supplemental Analysis

To shed further light on our theory, we conduct supplemental analysis to examine whether the interaction effect of *Supervisor Type* and *Performance Evaluation System* on employees' *Willingness to Engage in Extra-Role Behaviors* is serially mediated by *Quantification* and *Expected Reciprocity*. Following Newman et al. (2020), we measure *Quantification* using three items. Specifically, participants indicated on a seven-point scale (1 = *Strongly disagree*, 7 = *Strongly agree*) the extent to which they agreed with the following

statements: (1) “I feel like the evaluation process would just reduce me to a number,” (2) “the evaluation process would have trouble adequately recognizing my qualitative attributes, abilities, and performance,” and (3) “I think some information about my performance would be lost in the evaluation process.” We conduct a reliability analysis on these three items and confirm high internal consistency (Cronbach’s alpha of 0.93), indicating that the items sufficiently capture a single underlying construct. We average participants’ responses across the three items into a single measure (*Quantification*), with higher values indicating that participants perceive greater quantification of the qualitative aspects of their performance.

We also measure *Expected Reciprocity* from the human supervisor (or the AI system) using two items, each on a seven-point scale (1 = *Not at all*, 7 = *A lot*). Specifically, participants were asked to indicate: (1) the extent to which they trust that their supervisor (or the AI system) will recognize any additional tasks they perform for the company, and (2) the extent to which they trust that their supervisor (or the AI system) will compensate them for any additional tasks that they perform for the company. We again conduct a reliability analysis on these two items and confirm high internal consistency (Cronbach’s alpha of 0.93), indicating that the items sufficiently capture a single underlying construct. Hence, we average participants' responses across the two items into a single measure (*Expected Reciprocity*), with higher values indicating that participants expect greater reciprocity from their supervisor.

We test for moderated serial mediation using PROCESS Model 85 (Hayes, 2022). *Supervisor Type* is the independent variable and is coded as 1 (0) in the AI (Human) supervisor condition. *Performance Evaluation System* is the moderator and is coded as 1 (0) in the Mixed (Objective) evaluation system condition. *Quantification* and *Expected Reciprocity* are as defined above and act as mediators, and *Willingness to Engage in Extra-Role Behaviors* serves as the

dependent variable. Table 2, Panel A presents the observed path coefficients. We find that *Supervisor Type* has a significantly positive effect on *Quantification* ($\beta = 0.82$, $p < 0.01$), indicating that subordinates perceive AI supervisors to quantify the qualitative aspects of their performance more relative to human supervisors. Next, we find that *Quantification* negatively affects *Expected Reciprocity* ($\beta = -0.60$, $p < 0.01$), indicating that subordinates who believe that their supervisor will quantify the qualitative aspects of their performance will in turn expect less reciprocity from their supervisor for their extra-role contributions. Finally, *Expected Reciprocity* positively affects subordinates' *Willingness to Engage in Extra-Role Behaviors* ($\beta = 0.18$, $p = 0.01$), indicating that subordinates are more likely to engage in extra-role behaviors when they expect that their supervisors will reciprocate (i.e., recognize and reward them for) their extra-role contributions. Table 2, Panel B shows that the indirect effects of supervisor type on subordinates' willingness to engage in extra-role behaviors are negative and significant, both when the performance evaluation system is purely objective (indirect effect = -0.09, 95% CI = [-0.22, -0.01]) and when the performance evaluation system is mixed (indirect effect = -0.30, 95% CI = [-0.62, -0.04]), only when both of our mediators (i.e., *Quantification* and *Expected Reciprocity*) are included in the model. In contrast, the indirect effects of supervisor type on subordinates' willingness to engage in extra-role behaviors are not significant, in either the objective or mixed condition, when only one of our mediators is included in the model. Table 2, Panel C reports that the magnitude of the serial mediation effect of *Supervisor Type* on *Willingness to Engage in Extra-Role Behaviors* via *Quantification* and *Expected Reciprocity* is significantly stronger in the *Mixed Evaluation System* condition than in the *Objective Evaluation System* condition (index of moderated mediation = −0.21, 95% CI = [−0.46, −0.03]). Furthermore, when *Quantification* and *Expected Reciprocity* are accounted for as serial mediators, the direct effect of *Supervisor Type*

on *Willingness to Engage in Extra-Role Behaviors* is no longer significant (p = 0.36), which suggests full mediation via *Quantification* and *Expected Reciprocity.* Taken together, these results support our theory that AI supervision reduces subordinates' willingness to engage in extra-role behaviors through a sequential process in which AI supervisors heighten subordinates' perceptions of quantification, which in turn reduces subordinates' expected reciprocity from their supervisor, with this process being more pronounced when subordinates are evaluated under a mixed performance evaluation system.

[Insert Table 2 about here]

4. Experiment 2: Method and Results

4.1. Method

The results of our first experiment demonstrate that subordinates with AI supervisors are less willing to engage in extra-role behaviors than those with human supervisors when the performance evaluation system contains both objective and subjective criteria. Building on this finding, we examine in a second experiment whether humanizing AI supervisors (i.e., by assigning them a name) can mitigate the negative effect of AI supervision on subordinates' willingness to engage in extra-role behaviors. Our second experiment utilizes a 2 × 2 between-subjects design.[9] As in Experiment 1, we manipulate *Supervisor Type* by informing participants that their supervisor is either a human or an AI system. We also manipulate the *Level of Humanization* by either providing or not providing the name of the supervisor (Hodge et al., 2021). When a name is provided, the supervisor is referred to as "Robin." We chose this name because Robin is considered gender-neutral and we did not want perceptions associated with the

[9] This experiment received Institutional Review Board approval at the authors' institutions.

supervisor's gender to confound our analysis (Pethig and Kroenung, 2023; Maas and Torres-González, 2011). Appendix C depicts how the human and AI supervisors are named.[10]

Another important design choice in Experiment 2 is that we hold performance evaluation system constant by using the *Mixed Evaluation System* condition across all conditions, given that (1) the strong negative effect of AI supervision on subordinates' willingness to engage in extra-role behaviors persists only in the *Mixed Evaluation System* condition and (2) evaluation systems that contain objective and subjective criteria are more prevalent in practice. We recruited 200 participants from Prolific using the same selection criteria as that of Experiment 1. The resulting demographic characteristics of our sample are similar to those reported for Experiment 1.[11] All other procedures and solicited variables are consistent with those of Experiment 1.

4.2. Results

4.2.1. Manipulation Checks

Similar to Experiment 1, we assess the effectiveness of our *Supervisor Type* manipulation by asking participants to identify which of the following was true in their scenario: "My supervisor was a person" or "My supervisor was an AI system" (Hodge et al., 2021). To test the effectiveness of our *Level of Humanization* manipulation, we asked participants: "Was Robin the name of your supervisor, or were you not given a name for your supervisor?" (Hodge et al., 2021). Participants selected either "Robin was the name of my supervisor" or "I was not given a name for my supervisor." We find that 188 out of the 200 recruited participants correctly answered *both* manipulation check questions, which indicates that our manipulations of

[10] The conditions in which the human and AI supervisors are unnamed are the same as that depicted in Appendix A.
[11] On average, the participants had 19 years of work experience. 95 percent indicated that they had received some type of performance evaluation at work. 54 percent of participants were male, and the mean age was 41 years. Each participant took an average of 7.53 minutes to complete the instrument and received $2.00 for completing the task, resulting in an average hourly wage of $15.93.

Supervisor Type and *Level of Humanization* appear to be successful. To mitigate noise induced by low quality responses (Bentley et al., 2025), we exclude those who did not correctly answer both manipulation check questions from subsequent analyses.

As in Experiment 1, we also administered an attention check question to verify that participants understood the case materials. Again, participants were asked whether their supervisor (or the AI system) evaluates their performance *solely* based on the predefined objectively measurable criteria (Yes or No). We find that thirty-one (out of 188) participants failed to correctly answer this question, indicating that they did not carefully attend to or adequately comprehend our experimental materials.[12] Therefore, we exclude these participants from subsequent analyses, yielding a final sample of 157 participants for Experiment 2.

4.2.2. Test of H2

H2 predicts that, under a mixed performance evaluation system, naming an AI supervisor will increase subordinates' willingness to engage in extra-role behaviors. Table 3, Panel A presents cell sizes, means, and standard deviations for *Willingness to Engage in Extra-Role Behaviors* across the four experimental conditions, and Figure 3 graphically depicts the pattern of cell means by condition. The pattern of cell means observed in Figure 3 is visually consistent with our predicted pattern in Figure 1, Panel B. Table 3, Panel B presents the ANOVA analysis with *Willingness to Engage in Extra-Role Behaviors* as the dependent variable, and Table 3, Panel C presents the results of simple effects tests. We document a significant main effect of *Supervisor Type* ($p < 0.01$), indicating that participants with an AI supervisor reported lower

[12] Of the 200 recruited participants, a total of 36 participants failed to correctly answer the attention check question. Hence, there appears to be significant overlap in our sample between those who failed to correctly answer the attention check question *and* failed to correctly answer both manipulation check questions. Thus, the responses from these participants are likely to be of lower quality (Bentley et al., 2025), which further supports our decision to exclude them from subsequent analyses.

willingness to engage in extra-role behaviors overall than participants with a human supervisor. Notably, this difference is significant only when supervisors are not given a name (4.46 versus 5.23, $p < 0.01$), which replicates our earlier findings from Experiment 1. Consistent with H2, we find that naming an AI supervisor significantly increases participants' willingness to engage in extra-role behaviors (4.92 versus 4.46, $p = 0.04$, one-tailed). In contrast, we do not observe a significant effect of naming a human supervisor on participants' willingness to engage in extra-role behaviors (5.23 versus 5.21, $p = 0.94$). Additionally, we no longer observe a significant effect of supervisor type on participants' willingness to engage in extra-role behaviors when supervisors are given a name (4.92 versus 5.21, $p = 0.25$). Taken together, these results support H2 and indicate that humanizing AI supervisors by giving them a name can serve to mitigate the negative impact of AI supervision on subordinates' willingness to engage in extra-role behaviors under mixed performance evaluation systems that contain both objective and subjective criteria.

[Insert Table 3 and Figure 3 about here]

5. Conclusion

In this study, we conduct two experiments to investigate how subordinates' willingness to engage in extra-role behaviors is jointly influenced by supervisor type and the firm's performance evaluation system. In our first experiment, we find that AI supervisors, as opposed to human supervisors, reduce subordinates' willingness to engage in extra-role behaviors to a greater extent when subordinates are evaluated under a mixed performance evaluation system that contains both objective and subjective criteria. Moreover, we find that this negative relationship between AI supervisors and extra-role behaviors is driven by subordinates' perception that AI supervisors will overly quantify their extra-role contributions, leaving less room for recognition and reciprocation during performance evaluations. In our second

experiment, we find that humanizing AI supervisors by giving them a name helps to mitigate the negative effect of AI supervisors on subordinates' willingness to engage in extra-role behaviors. This finding suggests that the humanization of AI supervisors may serve as an important and necessary strategy to offset some of the negative unintended consequences that AI supervisors have on subordinates' willingness to engage in extra-role behaviors, which impacts overall organizational welfare.

Our study is subject to several limitations that give rise to future research opportunities. Although naming represents one of the most common and practical ways to humanize technology, we acknowledge that there are alternative ways to humanize AI supervisors besides naming, e.g., assigning avatars to AI supervisors. Thus, we recognize that the results of our study may not generalize to other ways of humanizing AI supervisors. Hence, future research may find it fruitful to investigate how other humanization methods influence subordinates' perceptions of their AI supervisor. Relatedly, we intentionally chose a gender-neutral name for the AI supervisor in our study (Robin) in order to ensure a clean experimental design and control for the potential confounding effects of perceptions associated with the supervisor's gender on our results. However, supervisors' gender in performance evaluation contexts may act as an important contextual variable that fluctuates with subordinates' behavioral tendencies, such as their willingness to engage in extra-role behaviors, in the real-world. Hence, future research could examine whether and how more gendered names associated with AI supervisors affect subordinates' judgments and decisions.

Notwithstanding the above limitations, our study offers the following contributions to the management accounting literature and practice. We complement early studies in a relatively nascent literature stream that examines the effects of AI supervisors on subordinates' perceptions

and decision-making. Specifically, we document a significantly negative interaction effect of supervisor type and performance evaluation system on subordinates' willingness to partake in extra-role behaviors. Additionally, we demonstrate a mechanism that can be employed to attenuate this adverse effect associated with AI supervisors, namely, humanizing the AI supervisors. Against the backdrop of increased deployment of AI supervisors in practice (Lanz et al., 2024; Lee et al., 2015; Möhlmann et al., 2021; Soper, 2021; Wesche and Sonderegger, 2019), as well as the importance of extra-role behaviors to overall organizational success, our study helps organizations better understand how to leverage and integrate AI supervisors into the workplace and suggests that effective humanization is vital. While we focus primarily on extra-role behaviors, our theory and findings pertaining to AI supervision may also have implications on other outcome variables of interest in management accounting, such as effort or budget reporting.

References

Baek, T. H., & Kim, M. (2023). Ai robo-advisor anthropomorphism: The impact of anthropomorphic appeals and regulatory focus on investment behaviors. *Journal of Business Research*, 164, 114039.

Becton, J. B., Giles, W. F., & Schraeder, M. (2008). Evaluating and rewarding OCBs: Potential consequences of formally incorporating organizational citizenship behavior in performance appraisal and reward systems. *Employee Relations*, 30(5). 494-514.

Bentley, J. W., Bloomfield, M. J., Bloomfield, R. J., & Lambert, T. A. (2025). Drivers of public opinion on the acceptability of distorting performance measures. *The Accounting Review,* 100(1), 87-111.

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

Blau, P. M. (1964). *Exchange and Power in Social Life*. Wiley.

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

Bol, J. C., C. Brown, and L. LaViers. (2025). Context and design matter: Employee acceptance of using artificial intelligence to conduct performance evaluations. Working paper.

Bol, J. C., Fogel-Yaari, H., Grabner, I., & Sedatole, K. L. (2026). Promoting organizational citizenship behavior (OCB): Unintended gender disparities in the participation and evaluation of different types of OCB. *The Accounting Review*, 101(1), 27-55.

Bolino, M. C. (1999). Citizenship and impression management: Good soldiers or good actors? *The Academy of Management Review*, 24 (1), 82-98.

Bolino, M. C., Turnley, W. H., Gilstrap, J. B., & Suazo, M. M. (2010). Citizenship under pressure: What's a "good soldier" to do? *Journal of Organizational Behavior*, 31(6), 835-855.

Bordia, P., Restubog, S. L. D., & Tang, R. L. (2008). When employees strike back: Investigating mediating mechanisms between psychological contract breach and workplace deviance. *Journal of Applied Psychology*, 93(5), 1104-1117.

Brown, J. L., J. Burke, and A. K. Sauciuc. (2026). Using artificial intelligence to evaluate employees: the effects on recruitment, effort, and retention. *Journal of Management Accounting Research*, 38(1), 1-26.

Burney, L. C., Henle, C., & Widener, S. (2009). A path model examining the relations among strategic performance measurement system characteristics, organizational justice, and extra- and in-role performance. *Accounting, Organizations and Society*, 34(3-4), 305-341.

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

Cha, Y. (2025). Employees' reactions to algorithmic performance evaluation: Threats of evaluation bias and objectivity. *Journal of Information Systems*, 39(2), 1-20.

Chaminade, T., Hodgins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience*, 2(3), 206-216.

Chen, C. X., Hudgins, R., & Wright, W. F. (2022). The effect of advice valence on the perceived credibility of data analytics. *Journal of Management Accounting Research*, 34(2), 97-116.

Cheng, M. M., & Coyte, R. (2014). The effects of incentive subjectivity and strategy

communication on knowledge-sharing and extra-role behaviours. *Management Accounting Research*, 25, 119-130.
Cook, K. S., Cheshire, C., Rice, E. R. W., & Nakagawa, S. (2013). Social exchange theory. In DeLamater, J., Ward, A. (Eds.), *Handbook of social psychology*, 2nd edition, 61-88. New York: Springer.
Cropanzano, R., & Mitchell, M. S. (2005). Social exchange theory: An interdisciplinary review. *Journal of Management*, 31(6), 874-900.
Cui, H., & Yasseri, T. (2026). Gender bias in perception of human managers extends to AI managers. *Computers in Human Behavior Reports*, 21, 100984.
Datar, S., Kulp, S., & Lambert, R. (2001). Balancing performance measures. *Journal of Accounting Research*, 39(1), 75-92.
Drake, A. R., Haka, S. F., & Ravenscroft, S. P. (1999). Cost system and incentive structure effects on innovation, efficiency, and profitability in teams. *The Accounting Review*, 74, 323-345.
Eastman, K. K. (1994). In the eyes of the beholder: An attributional approach to ingratiation and organizational citizenship behavior. *Academy of Management Journal,* 37, 1379-1391.
El Halabi, M., & Trendel, O. (2024). Just name it: The act of naming humanoid service robots decreases perceived eeriness and increases repurchase intent. *Journal of Service Research*, 28(1), 131-149.
Feltham, G., & Xie, J. (1994). Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review*, 69, 429-453.
Ganbold, O., Rose, A. M., Rose, J. M., & Rotaru, K. (2022). Increasing reliance on financial advice with avatars: The effects of competence and complexity on algorithm aversion. *Journal of Information Systems*, 36(1), 7-17.
Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, 24(4), 1494-1509.
Harms, P. D., & Han, G. (2019). Algorithmic leadership: The future is now. *Journal of Leadership Studies*, 12(4), 74-75.
Hayes, A. F. (2022). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Guilford Press.
He, W., Li, S. L., Feng, J., Zhang, G., & Sturman, M. C. (2021). When does pay for performance motivate employee helping behavior? The contextual influence of performance subjectivity. *Academy of Management Journal*, 64(1), 293-326.
Höddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, 116, 106635.
Hodge, F. D., Mendoza, K. I., & Sinha, R. K. (2021). The effect of humanizing robo-advisors on investor judgments. *Contemporary Accounting Research*, 38(1), 770-792.
Holland, C., Levis, J., Nuggehalli, R., Santilli, B., & Winters, J. (2017). UPS optimizes delivery routes. *Interfaces*, 47(1), 8–23.
Johnson, L. (2017). Why tech companies like IBM and Amazon brand artificial intelligence with human names. Adweek. https://www.adweek.com/performance-marketing/why-tech-companies-like-ibm-and-amazon-brand-artificial-intelligence-with-human-names/
Kunz, J. (2015). Objectivity and subjectivity in performance evaluation and autonomous motivation: An exploratory study. *Management Accounting Research*, 27, 27-46.
Lanz, L., Briker, R., & Gerpott, F. H. (2024). Employees adhere more to unethical instructions

from human than AI supervisors: Complementing experimental evidence with machine learning. *Journal of Business Ethics*, 189, 625-646.
Lecher, C. (2019). How Amazon automatically tracks and fires warehouse workers for 'productivity.' The Verge. https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centers-productivity-firing-terminations (accessed 23 June 2026)
Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 1-16.
Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. *CH'15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1603-1612.
Lefkowitz, J. (2000). The role of interpersonal affective regard in supervisory performance ratings: A literature review and proposed causal model. *Journal of Occupational and Organizational Psychology*, 7, 67-85.
Lin, F., Cheng, M. M., & Humphreys, K. A. (2026). Tough ratings, tougher sell: How different types of adjustment affect managers' asymmetric algorithm use in performance evaluation judgments. *The Accounting Review*, 101(3), 413-440.
Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90-103.
Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, 46(4), 629–650.
Maas, V., & Torres-González, R. (2011). Subjective performance evaluation and gender discrimination. *Journal of Business Ethics*, 101(4), 667-681.
Marinova, S. V., Cao, X., & Park, H. (2018). Constructive organizational values climate and organizational citizenship behaviors: A configurational view. *Journal of Management*, 45(5), 2045-2071.
Marinova, S. V., Moon, H., & Kamdar, D. (2013). Getting ahead or getting along? The two-facet conceptualization of conscientiousness and leadership emergence. *Organization Science*, 24(4), 1257-1276.
Möhlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. W. (2021). Algorithmic management of work on online labor platforms: When matching meets control. *MIS Quarterly*, 45(4), 1999-2022.
Nan, X. Anghelcev, G., Myers, J. R., Sar, S., & Faber, R. (2006). What if a web site can talk? Exploring the persuasive effects of web-based anthropomorphic agents. *Journalism & Mass Communication Quarterly*, 83(3), 615-631.
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.
Nunnally, J. (1978). *Psychometric Theory*. McGraw-Hill.
Organ, D. W. (1988). *Organizational Citizenship Behavior: The Good Solider Syndrome*. Lexington Books.
Organ, D. W., & Ryan K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology*, 48(4), 775-802.
Parent-Rocheleau, X., & Parker, S. K. (2022). Algorithms as work designers: How algorithmic management influences the design of jobs. *Human Resource Management Review*, 32(3), 100838.

Pethig, F., & Kroenung, J. (2023). Biased humans, (un)biased algorithms? *Journal of Business Ethics*, 183, 637-652.
Pickard, M. D., Schuetzler, R., Valacich, J. S., & Wood, D. A. (2020). Innovative accounting interviewing: A comparison of real and virtual accounting interviewers. *The Accounting Review*, 95(6), 339-366.
Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M., & Blume, B. D. (2009). Individual- and organizational-level consequences of organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology*, 94(1), 122-141.
Podsakoff, P. M., Ahearne, M., & MacKenzie, S. B. (1997). Organizational citizenship behavior and the quantity and quality of work group performance. *Journal of Applied Psychology*, 82(2), 262-270.
Podsakoff, P. M., MacKenzie, S. B., & Hui, C. (1993). Organizational citizenship behaviors and managerial evaluations of employee performance: A review and suggestions for future research. In G. R. Ferris (Ed.), *Research in personnel and human resources management,* vol. 11, 1-40. Greenwich, CT: JAI Press.
Podsakoff, P. M., MacKenzie, S. B., Paine, J., & Bachrach, D. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management*, 26(3), 513-563.
Qiao, Y., & Song, D. (2025). The effect of robot anthropomorphism on customers' attention activation and satisfaction. *Journal of Consumer Behavior: An International Research Review*, 24(1), 201-219.
Restubog, S. L. D., Hornsey, M., Bordia, P. & Esposo, S. (2008). Effects of psychological contract breach on organizational citizenship behavior: Insights from the group value model. *Journal of Management Studies*, 45(8), 1377-1399.
Resume Builder. (2025). Half of managers use AI to determine who gets promoted and fired. https://www.resumebuilder.com/half-of-managers-use-ai-to-determine-who-gets-promoted-and-fired/ (accessed 23 June 2026)
Singla, A., Sukharevsky, A., Yee, L., Chui, M., & Hall, B. (2024). The state of AI in early 2024: Gen AI adoption spikes and starts to generate value. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024 (accessed 23 June 2026)
Soper, S. (2021). Fired by bot at Amazon: 'It's you against the machine.' Bloomberg. https://www.bloomberg.com/news/features/2021-06-28/fired-by-bot- amazon-turns-to-machine-managers-and-workers-are-losing-out
Tabansi, C. (2025). How UPS' ORION algorithm transformed its route optimization. Supply Chain Nuggets. https://supplychainnuggets.com/how-ups-orion-algorithm-transformed-its-route-optimization/ (accessed 23 June 2026)
Taylor, E. Z. (2006). The effect of incentives on knowledge sharing in computer-mediated communication: An experimental investigation. *Journal of Information Systems*, 20(1), 103-116.
Thibaut, J. W., & Kelley, H. H. (1959). *The Social Psychology of Groups*. Wiley.
Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52, 113-117.
Werner, J. M. (1994). Dimensions that make a difference: Examining the impact of in-role and extra-role behaviors on supervisory ratings. *Journal of Applied Psychology*, 79, 98-107.

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

Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, 17(3), 601-617.

Zhou, L., & Wang, V. L. (2025). "What's in a name?": The effect of AI agent naming on psychological ownership and responsible behaviors in the shared economy. *Journal of Applied Business & Behavioral Sciences*, 1(2), 145-163.

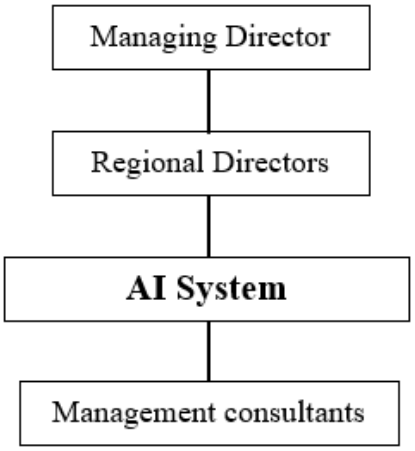
Appendix A
***Supervisor Type* Manipulation**

Panel A: *Human Supervisor* Condition



Your supervisor is a person who serves as a link between management consultants and upper-level management. Your supervisor evaluates your performance and provides feedback to you.

Panel B: *AI Supervisor* Condition



Your supervisor is an artificial intelligence (AI) system that serves as a link between management consultants and upper-level management. The AI system evaluates your performance and provides feedback to you.

Appendix B
Performance Evaluation System Manipulation

Panel A: *Objective Evaluation System* Condition

Your performance as a management consultant at Rivion Consulting is evaluated by **your supervisor (or the AI system)** solely based on the following four predefined performance criteria:

- Billable hours
- Average billing rate you charged to your clients this month ($)
- Total revenue from new project work requested by your clients this month
- Evaluation ratings by your clients

Your supervisor (or The AI system) evaluates your work **exclusively** using these **objectively measurable criteria**. The performance evaluation **does *NOT* involve subjective judgment**. The final evaluation outcome is therefore determined **entirely by these predefined measurable criteria.**

Your salary and bonus depend to a significant degree on the outcome of your performance evaluation.

Panel B: *Mixed Evaluation System* Condition

Your performance as a management consultant at Rivion Consulting is evaluated by **your supervisor (or the AI system)** based in part on the following four predefined performance criteria:

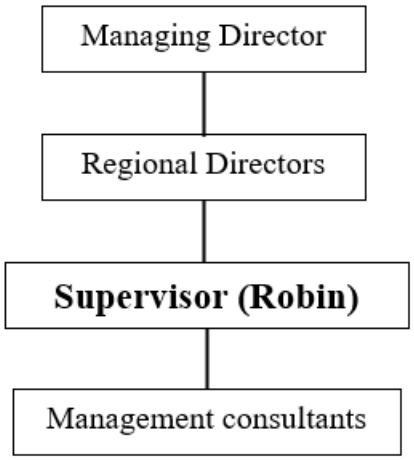
- Billable hours
- Average billing rate you charged to your clients this month ($)
- Total revenue from new project work requested by your clients this month
- Evaluation ratings by your clients

In addition to these objectively measurable criteria, your supervisor (or the AI system) also evaluates your work using **subjective criteria** (e.g., leadership skills, contributions to teamwork, potential for growth, etc.). The performance evaluation therefore **involves subjective judgment**. The final evaluation outcome is therefore determined **by both the predefined measurable criteria and additional judgement-based considerations.**

Your salary and bonus depend to a significant degree on the outcome of your performance evaluation.

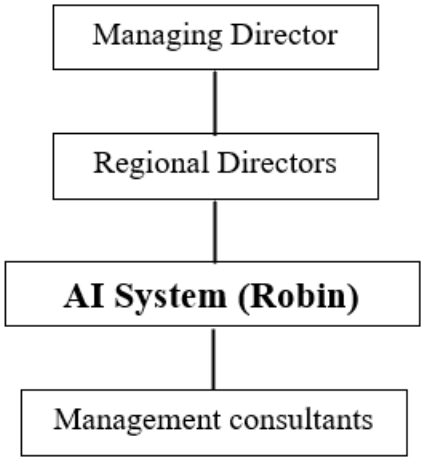
Appendix C
Named Supervisor Conditions

Panel A: Human Supervisors with a Name



Your supervisor is Robin, a person who serves as a link between management consultants and upper-level management. Robin evaluates your performance and provides feedback to you.

Panel B: AI Supervisors with a Name



Your supervisor is Robin, an artificial intelligence (AI) system that serves as a link between management consultants and upper-level management. Robin evaluates your performance and provides feedback to you.

Table 1
The Effect of Supervisor Type and Performance Evaluation System on Subordinates' Willingness to Engage in Extra-Role Behaviors

Panel A: Descriptive Statistics of *Willingness to Engage in Extra-Role Behaviors* by Condition

	Performance Evaluation System		
Supervisor Type	Objective	Mixed	Row Total
Human	4.42	5.52	4.98
	(1.42)	(0.91)	(1.31)
	n=46	n=47	n=93
AI	4.45	4.78	4.61
	(1.43)	(1.57)	(1.50)
	n=49	n=44	n=93
Column Total	4.44	5.16	4.79
	(1.42)	(1.32)	(1.41)
	n=95	n=91	n=186

Panel B: ANOVA Analysis of *Willingness to Engage in Extra-Role Behaviors*

Source of Variation	**SS**	**df**	**MS**	**F-stat**	**p-value**
Supervisor Type	5.90	1	5.90	3.21	0.07
Performance Evaluation System	23.90	1	23.90	13.07	<0.01
Supervisor Type × Performance Evaluation System	6.90	1	6.90	3.75	0.05
Error	332.70	182	1.80		

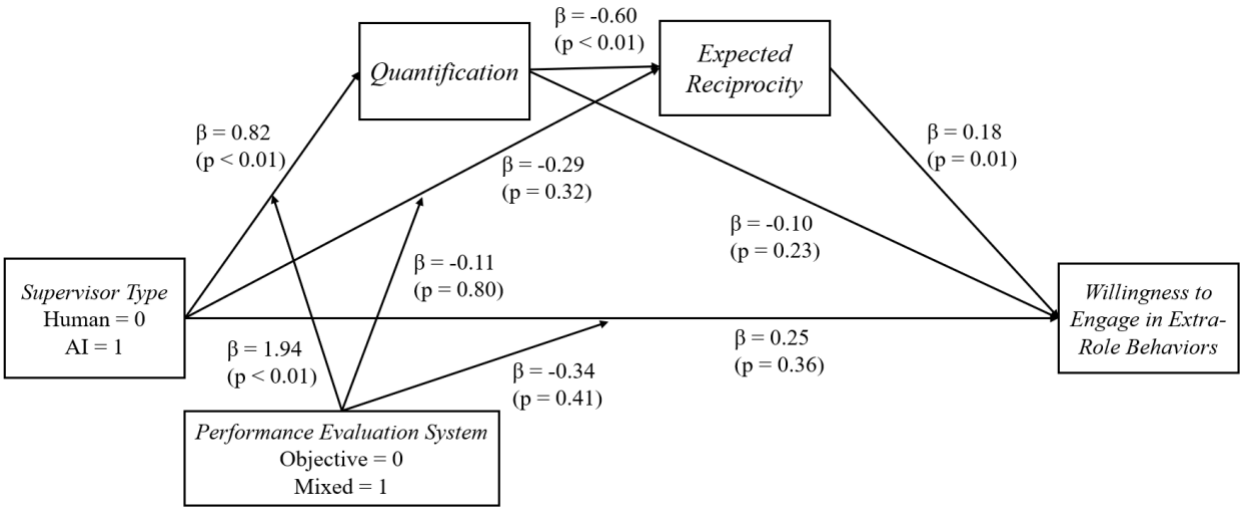
Panel C: Tests of Simple Effects for *Willingness to Engage in Extra-Role Behaviors*

Source of Variation	**df**	**t-stat**	**p-value**
Effect of Supervisor Type given Objective Evaluation System condition	1	0.10	0.92
Effect of Supervisor Type given Mixed Evaluation System condition	1	2.62	0.01
Effect of Performance Evaluation System given Human Supervisor condition	1	3.93	<0.01
Effect of Performance Evaluation System given AI Supervisor condition	1	1.19	0.24

Notes: *Willingness to Engage in Extra-Role Behavior* is computed as the average of participants' willingness to devote time to each of the following five tasks: (1) "help others who have been absent," (2) "help others who have heavy workloads," (3) "listen to co-workers' problems and worries," (4) "take a personal interest in other employees," and (5) "pass along information to co-workers." Participants' responses to each of these statements take on a value ranging from 1 ("Strongly disagree") to 7 ("Strongly agree"). Standard deviations are reported in parentheses.

Table 2
Moderated Serial Mediation Analysis

Panel A: Observed Path Coefficients



Panel B: Indirect Effects and Confidence Intervals

Mediator(s)	Performance Evaluation System	Effect	95% Confidence Interval
Quantification only	Objective	-0.08	[-0.25, 0.04]
	Mixed	-0.28	[-0.77, 0.13]
Expected Reciprocity only	Objective	-0.05	[-0.20, 0.08]
	Mixed	-0.07	[-0.27, 0.05]
Both *Quantification* and *Expected Reciprocity*	Objective	-0.09	[-0.22, -0.01]
	Mixed	-0.30	[-0.62, -0.04]

Panel C: Index of Moderated Mediation

Index	Bootstrapped Standard Error	95% Confidence Interval
-0.21	0.11	[-0.46, -0.03]

Notes: This table presents the results of moderated serial mediation analysis using Model 85 of the Hayes (2022) PROCESS macro. *Supervisor Type* is coded as 1 (0) in the *AI* (*Human*) *Supervisor* condition. *Performance Evaluation System* is coded as 1 (0) in the *Mixed* (*Objective*) *Evaluation System* condition. *Willingness to Engage in Extra-Role Behavior* is computed as the average of participants' responses to their willingness to devote time to each of the following five tasks: (1) "help others who have been absent," (2) "help others who have heavy workloads," (3) "listen to co-workers' problems and worries," (4) "take a personal interest in other employees," and (5) "pass along information to co-workers." Participants' responses to each of these statements take on a value ranging from 1 ("Strongly disagree") to 7 ("Strongly agree"). Quantification is computed as the average of participants' degree of agreement to the following three items: (1) "I feel like the evaluation would just reduce me to a number," (2) The evaluation process would have trouble adequately recognizing my qualitative attributes, and abilities, and performance," and (3) I think some information about my performance would be lost in the evaluation process. Participants' response to each of these statements take on a value ranging from 1 ("Strongly disagree") to 7

("Strongly agree"). *Expected Reciprocity* is computed as the average of participants' responses to the following two questions: (1) "To what extent do you trust that your supervisor (or the AI system) will recognize any additional tasks that you perform for the company," and (2) "To what extent do you trust that your supervisor (or the AI system) will compensate you for any additional tasks that you perform for the company." Participants' responses to each of these questions take on a value ranging from 1 ("Not at all") to 7 ("A lot").

Table 3
The Impact of Supervisor Type and Level of Humanization on Subordinates' Willingness to Engage in Extra-Role Behaviors

Panel A: Descriptive Statistics of *Willingness to Engage in Extra-Role Behaviors* by Condition

	Level of Humanization		
Supervisor Type	Low (Unnamed Supervisor)	High (Named Supervisor)	Row Total
Human	5.23	5.21	5.22
	(0.95)	(0.96)	(0.95)
	n=43	n=43	n=86
AI	4.46	4.92	4.68
	(1.43)	(1.05)	(1.28)
	n=37	n=34	n=71
Column Total	4.87	5.08	4.97
	(1.25)	(1.00)	(1.14)
	n=80	n=77	n=157

Panel B: ANOVA Analysis of *Willingness to Engage in Extra-Role Behaviors*

Source of Variation	**SS**	**df**	**MS**	**F-stat**	**p-value**
Supervisor Type	11.00	1	11.00	8.95	<0.01
Level of Humanization	1.90	1	1.90	1.54	0.22
Supervisor Type × Level of Humanization	2.20	1	2.20	1.81	0.18
Error	186.70	153	1.20		

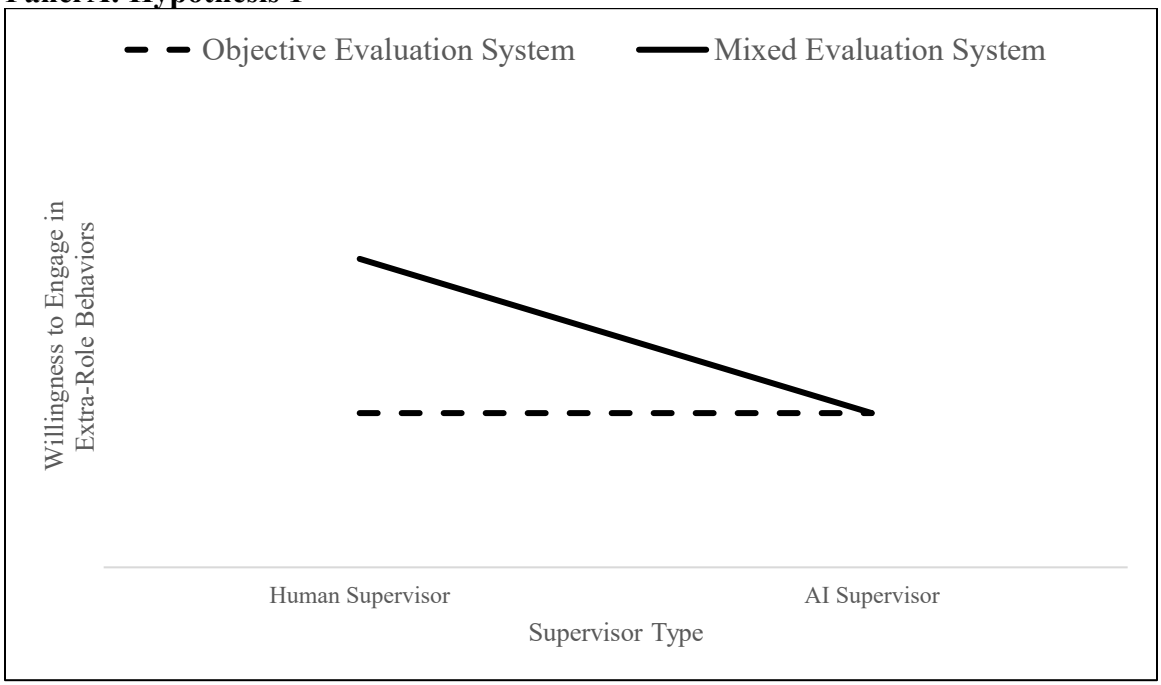
Panel C: Tests of Simple Effects for *Willingness to Engage in Extra-Role Behaviors*

Source of Variation	**df**	**t-stat**	**p-value**
Effect of Supervisor Type given Low Humanization condition	1	3.10	<0.01
Effect of Supervisor Type given High Humanization condition	1	1.15	0.25
Effect of Level of Humanization given Human Supervisor condition	1	0.10	0.94
Effect of Level of Humanization given AI Supervisor condition	1	1.75	0.04[a]

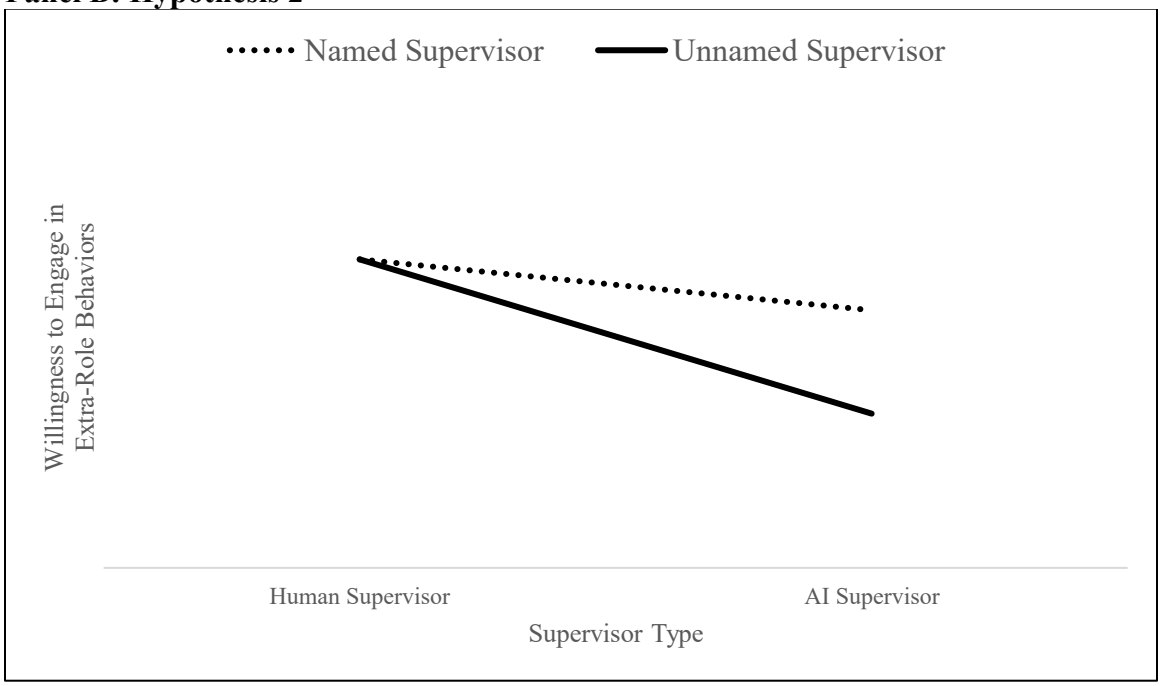
Notes: *Willingness to Engage in Extra-Role Behavior* is computed as the average of participants' willingness to devote time to each of the following five tasks: (1) "help others who have been absent," (2) "help others who have heavy workloads," (3) "listen to co-workers' problems and worries," (4) "take a personal interest in other employees," and (5) "pass along information to co-workers." Participants' responses to each of these statements take on a value ranging from 1 ("Strongly disagree") to 7 ("Strongly agree"). Standard deviations are reported in parentheses. [a]p-value is one-tailed equivalent for directional prediction.

Figure 1
Pattern of Predictions

Panel A: Hypothesis 1



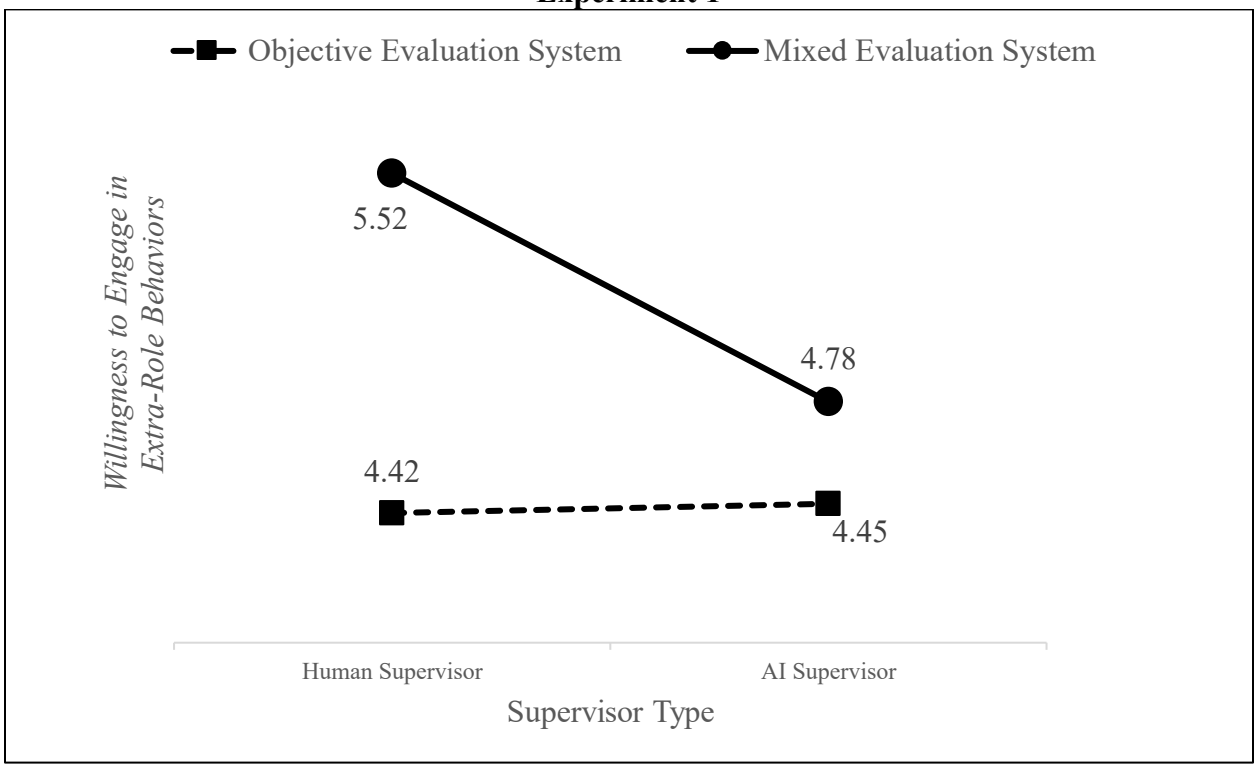
Panel B: Hypothesis 2



Notes: This figure depicts the predicted pattern of results for H1 (Panel A) and H2 (Panel B). H1: Subordinates' willingness to engage in extra-role behaviors is lower under AI supervision than under human supervision when the performance evaluation system contains subjective criteria, but supervisor type does not influence subordinates' willingness to engage in extra-role behaviors when the performance evaluation system contains purely objective

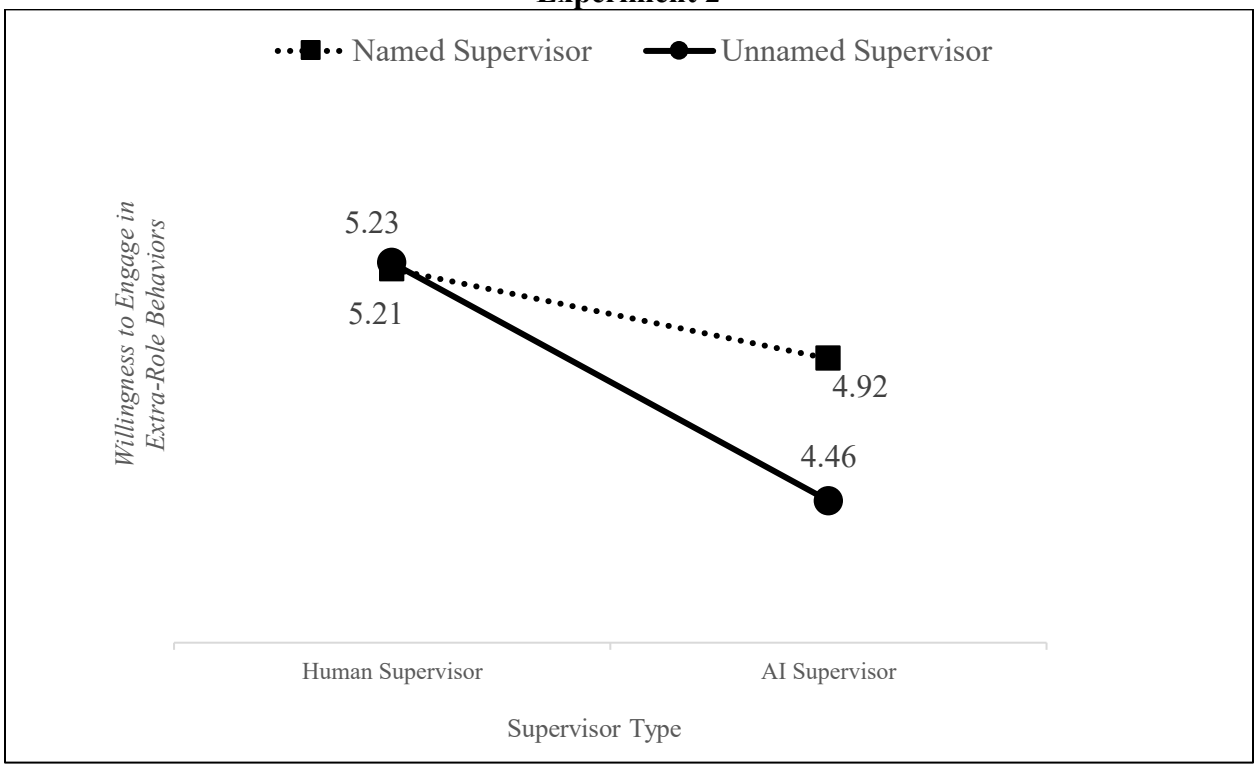
criteria. H2: Humanizing an AI supervisor by giving it a name will *increase* subordinates' willingness to engage in extra-role behaviors under a mixed performance evaluation system that contains subjective criteria.

Figure 2
Observed Means of *Willingness to Engage in Extra-Role Behaviors* by Condition – Experiment 1



Notes: This figure depicts the mean of *Willingness to Engage in Extra-Role Behavior* by experimental condition in Experiment 1. *Willingness to Engage in Extra-Role Behavior* is computed as the average of participants' willingness to devote time to each of the following five tasks: (1) "help others who have been absent," (2) "help others who have heavy workloads," (3) "listen to co-workers' problems and worries," (4) "take a personal interest in other employees," and (5) "pass along information to co-workers." Participants' responses to each of these statements take on a value ranging from 1 ("Strongly disagree") to 7 ("Strongly agree").

Figure 3
Observed Means of *Willingness to Engage in Extra-Role Behaviors* by Condition – Experiment 2



Notes: This figure depicts the mean of *Willingness to Engage in Extra-Role Behavior* by experimental condition in Experiment 2. *Willingness to Engage in Extra-Role Behavior* is computed as the average of participants' willingness to devote time to each of the following five tasks: (1) "help others who have been absent," (2) "help others who have heavy workloads," (3) "listen to co-workers' problems and worries," (4) "take a personal interest in other employees," and (5) "pass along information to co-workers." Participants' responses to each of these statements take on a value ranging from 1 ("Strongly disagree") to 7 ("Strongly agree").