Check for updates

Journal of Organizational Behavior

WILEY

Journal of Organizational Behavior

RESEARCH ARTICLE OPEN ACCESS

Is Our Future Colleague Even Human? Advancing Human-AI Teamwork from An Organizational Perspective

# Navigating AI Convergence in Human–Artificial Intelligence Teams: A Signaling Theory Approach

Andria Smith[1] | Hunter Phoenix van Wagoner[2] | Ksenia Keplinger[1] | Can Celebi[3]

[1]Independent Research Group "Organizational Leadership and Diversity", Max Planck Institute for Intelligent Systems, Stuttgart, Germany | [2]Department of Management, California State University Fullerton, Fullerton, California, USA | [3]Vienna Center for Experimental Economics, University of Vienna, Vienna, Austria

**Correspondence:** Ksenia Keplinger (kkeplinger@is.mpg.de)

**Received:** 5 December 2023 | **Revised:** 19 November 2024 | **Accepted:** 17 December 2024

**Funding:** This work was supported by Joachim Herz Stiftung and Max Planck Society.

**Keywords:** AI convergence | artificial intelligence | human–AI teaming | optional advice | signaling theory

## ABSTRACT

Teams that combine human intelligence with artificial intelligence (AI) have become indispensable for solving complex tasks in various decision-making contexts in modern organizations. However, the factors that contribute to AI convergence, where human team members align their decisions with those of their AI counterparts, still remain unclear. This study integrates signaling theory with self-determination theory to investigate how specific signals—such as signal fit, optional AI advice, and signal set congruence—affect employees' AI convergence in human–AI teams. Based on four experimental studies conducted in facial recognition and hiring contexts with approximately 1100 participants, the findings highlight the significant positive impact of congruent signals from both human and AI team members on AI convergence. Moreover, providing an option for employees to solicit AI advice also enhances AI convergence; when AI signals are chosen by employees rather than forced upon them, participants are more likely to accept AI advice. This research advances knowledge on human–AI teaming by (1) expanding signaling theory into the human–AI team context; (2) developing a deeper understanding of AI convergence and its drivers in human–AI teams; (3) providing actionable insights for designing teams and tasks to optimize decision-making in high-stakes, uncertain environments; and (4) introducing facial recognition as an innovative context for human–AI teaming.

## 1 | Introduction

In a transformative era of artificial intelligence (AI) integration, studying AI convergence, where human team members align their decisions with those of their AI counterparts, can be of immense benefit to organizations. AI systems have emerged as accelerators in the workplace, offering exceptional convenience and transforming the way complex tasks are performed and the speed at which decisions are made (Howard, Rabbitt, and Sirotin 2020). Advancements in AI have reshaped the status quo across various contexts, from monitoring remote work environments and making decisions in criminal justice and finance to influencing healthcare outcomes and processes for hiring and promoting employees (Kaushik 2022). Today, AI systems are increasingly seen as collaborative teammates for human workers, with human–AI teaming focusing on interactions where both work together to solve collaborative tasks and achieve shared goals (Zhao, Simmons, and Admoni 2022). Understanding when, how, and why human–AI teams converge in solving complex tasks

Andria Smith and Hunter Phoenix van Wagoner share the first authorship.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Journal of Organizational Behavior* published by John Wiley & Sons Ltd.

under high uncertainty is crucial for promoting effective decision-making processes that satisfy the needs for autonomy and competence of human employees (Gagné et al. 2022; Vaccaro, Almaatouq, and Malone 2024).

Human–AI teams, in which the AI system augments one or more employees, are meant to synthesize the best of human intelligence and AI (Bansal et al. 2021). However, beneath their apparent technological prowess lies a complex challenge—the interplay between AI advice and the factors that influence the acceptance or rejection of such advice. Applying signaling theory (Connelly et al. 2011; Drover, Wood, and Corbett 2018) to the context of human–AI teaming and integrating it with self-determination theory (Deci and Ryan 1985), we investigate AI convergence across different domains that directly impact people's work, safety, and careers. Specifically, we focus on domains characterized by high uncertainty and high importance, such as facial recognition and hiring, highlighting the responsibility of human leaders for decisions made by human–AI teams.

Human–AI teams make for a promising application of signaling theory for three reasons. First, in human–AI teams, AI advice is derived from processing vast amounts of data and performing complex analyses, often beyond employees' comprehension (Mosqueira-Rey et al. 2022). Second, AI acts as a signal sender to human teammates, with traits based on data-driven models that may contain unnoticed imperfections (Ali et al. 2023). Third, signaling theory offers a framework to interpret both AI signals and human signals, as well as the congruence between them in different contexts.

Our research explores how three key aspects of signaling theory—signal fit, optional AI advice, and congruence between human and AI signals—affect AI convergence in human–AI teams. We designed a series of experiments involving either a facial recognition task (Studies 1–3) or a CV screening task (Study 4) to manipulate signal fit (varying information about the credibility of AI advice), optional versus mandatory AI advice, and congruent versus incongruent signals from both human and AI senders. Across four experimental studies involving over 1100 participants, we tested our hypotheses to shed light on the main factors influencing AI convergence and provide actionable insights for organizations aiming to foster effective human–AI teams in the workplace.

Our study contributes to existing knowledge in four ways. First, we extend signaling theory by integrating it with self-determination theory and applying this integrated framework to the context of human–AI teams. By examining specific signal characteristics (e.g., observability and fit), multiple signal sources (human and AI), and various signal sets (congruent and incongruent), we uncover novel theoretical mechanisms that determine AI convergence within teams comprising both human and AI members. In addition, we apply signaling theory within teams, demonstrating how signals from both human and AI team members can impact decision-making processes internally. Although most studies apply signals as a method to communicate to outsiders, our research illustrates that signaling theory can also explain the behavior of internal, and not solely external, stakeholders. These insights can inform further theoretical development on optimizing team composition comprising multiple humans and AI and on understanding convergence among diverse signal sources.

Second, we follow recent calls to develop a better understanding of human–AI collaboration in organizational settings (Köchling and Wehner 2020; Raisch and Krakowski 2021; Vaccaro, Almaatouq, and Malone 2024). We contribute to the literature by developing a deeper understanding of AI convergence in human–AI teams and the factors influencing it, aiming to empower individuals to leverage AI's capabilities while fulfilling employees' need for autonomy and competence and preserving agency in decision-making processes. Our results lend weight to the augmentation thesis favoring a collaborative model of AI configuration in which humans and AI influence each other in iterative processes (Murray, Rhymer, and Sirmon 2021; Raisch and Krakowski 2021).

Third, our research contributes to a more comprehensive empirical understanding of human–AI teaming by developing impactful, actionable recommendations aimed at navigating AI convergence. The integration of human–AI teams into decision-making processes within highly complex and uncertain environments may give rise to ethical and privacy concerns, influencing employees' acceptance of AI advice due to the potentially significant consequences of their decisions. Understanding these implications and applying careful team and task design can impact AI convergence and optimize decision-making in high-stakes environments.

Fourth, although human–AI teams in the context of hiring have been extensively studied by researchers (e.g., van den Broek, Sergeeva, and Huysman 2021), the application of human–AI teaming in the context of facial recognition represents a novel area for exploration within organizational settings. Introducing this novel context to the field of organizational behavior is important and interesting because it presents unique challenges and ethical considerations regarding privacy, bias mitigation, and the integration of AI technologies into sensitive operational domains.

## 2 | Human–AI Teaming

The prevalence of employee interaction with AI is rapidly increasing, especially in modern workplaces. Today, AI tools transform organizational processes by supporting humans in executing managerial functions, trading on the stock market, and even making life-altering decisions, such as hiring and healthcare choices (Chugunova and Sele 2022; Gonzalez et al. 2022). Human–AI teaming involves humans and AI systems both collaborating and working interdependently to achieve shared objectives by leveraging their complementary strengths and weaknesses (Dellermann et al. 2019; Einola and Khoreva 2022; Hollenbeck, Beersma, and Schouten 2012; Memmert and Bittner 2022). Many AI systems are designed for collaborative settings, serving as advisory tools and forming human–AI teams. These systems are evolving beyond mere tools and are increasingly perceived as collaborative teammates, prompting the development of new forms of work and cooperation (Seeber et al. 2020; Wang, Maes, et al. 2021). Specifically, AI teammates are emerging as highly impactful contributors because of their



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

advanced knowledge processing capabilities, sensing abilities, and natural language interaction with human teammates (Seeber et al. 2020).

Although research on human–AI teaming in organizational settings is scattered across various fields, such as computer science, economics, psychology, and organizational behavior (Đula et al. 2023; Vaccaro, Almaatouq, and Malone 2024), researchers agree that employees react differently to AI teammates compared with fellow human teammates. In particular, when interacting with AI teammates, individuals tend to show less regard to social norms, demonstrate less emotions, and act with increased rationality compared with human-to-human interactions (Castelo 2019; Chugunova and Sele 2022; Jago 2019). Employee's reactions to AI teammates' advice vary widely, ranging from algorithm aversion to algorithm appreciation. Algorithm aversion refers to the situation when human team members exhibit a reluctance to engage with algorithms, favoring human teammates instead, even if algorithms outperform humans (Dietvorst, Simmons, and Massey 2015; Jussupow, Benbasat, and Heinzl 2020). For example, Gonzalez et al. (2022) observed that job candidates feel less confident and in control when AI systems handle hiring decisions without human input, as it limits their ability to showcase skills. Similarly, a study by Newman, Fast, and Harmon (2020) found that humans possess the capability to consider individual circumstances and foundational factors impacting employees' performance in a more comprehensive manner compared to algorithms. Algorithm aversion in the workplace may arise from employee fears (Bankins et al. 2024; Tong et al. 2021) regarding (a) the risk of job loss and adverse career impact (Suseno et al. 2022), (b) skills becoming redundant due to automation risk (Innocenti and Golin 2022), and (c) substantial alterations in existing jobs due to technological advancements (Brougham and Haar 2020).

However, there is empirical evidence that employees may also be grateful for AI advice and may even prefer it compared to human advice—a phenomenon termed algorithm appreciation (Logg, Minson, and Moore 2019). For example, in regard to investment decisions, Keding and Meissner (2021) found that in objective task settings, managers saw AI advice as more valid than identical human suggestions. Algorithm appreciation arises from the belief that algorithms employ a more systematic decision-making process, enhancing decision quality (Bankins et al. 2024). Furthermore, although bias in AI exists (e.g., Buolamwini and Gebru 2018), a recent study suggests that women favor algorithmic recruiters over male recruiters because of the anticipation of more favorable assessments from AI (Pethig and Kroenung 2022).

Accordingly, we argue that synergies in human–AI teams are not guaranteed. In their recent review, Vaccaro, Almaatouq, and Malone (2024) illustrate this point by highlighting the importance of context and content in optimal human–AI teams. Previous research suggests that there are three main factors influencing the likelihood of AI convergence in human–AI teams: decision context, agency in decision-making, and individual differences (Bankins et al. 2024; Chugunova and Sele 2022). For example, although employees seem to be more receptive to accepting AI advice in analytical or objective contexts, they prefer human advice in social or moral settings because they perceive AI to be unable to account for the uniqueness of humans (Fumagalli, Rezaei, and Salomons 2022; Haesevoets et al. 2021; Longoni, Bonezzi, and Morewedge 2019). In addition, individuals seem to rely on AI advice in situations of high personal importance (Saragih and Morrison 2022), high uncertainty (Altintas, Seidmann, and Gu 2023), and increasing task difficulty (Bogert, Schecter, and Watson 2021; Walter, Kremmel, and Jäger 2021). In terms of the distribution of agency in decision-making, although employees are hesitant to completely hand over control of their decisions, they are willing to rely on AI advice, when they retain the ultimate decision authority (Haesevoets et al. 2021). Dietvorst, Simmons, and Massey (2018) show that managers are willing to accept AI advice when they have majority control, with 70% human input and 30% AI input. Remarkably, AI systems that incorporate more social cues tend to positively influence individuals' preferences toward AI advice, eliciting responses similar to those typically observed in human–human interactions (Jussupow, Benbasat, and Heinzl 2024).

As for individual differences, when employees have high confidence in their own abilities to perform a task successfully, they are less likely to solicit and follow AI advice (Logg, Minson, and Moore 2019; Snijders et al. 2022). Similarly, when individuals have professional expertise in a particular decision-making domain, they are more inclined to disregard AI advice (Burton, Stein, and Jensen 2020; Thurman, Lewis, and Kunert 2019). In general, individuals exhibit high levels of trust in AI teammates, often following their advice even when provided with minimal information about how they operate (Gillath et al. 2021; Kennedy, Waggoner, and Ward 2021; You, Yang, and Li 2022). Knowledge about AI as well as experience in using it for completing tasks has been demonstrated to enhance the likelihood of AI convergence (Burton, Stein, and Jensen 2020; Chong et al. 2022; Yeomans et al. 2019).

Finally, the quality of AI advice is another crucial aspect of AI convergence. Employees are more inclined to accept AI advice that has proven highly accurate in the past (Sergeeva et al. 2023). Conversely, when an AI provides incorrect advice, it makes human teammates doubt the AI to a greater extent compared with when fellow human teammates make easily identifiable mistakes (Madhavan, Wiegmann, and Lacson 2006). Providing appropriate explanations for AI-assisted decision-making enhances AI convergence, regardless of the correctness of AI advice (Bansal et al. 2021; Lai and Tan 2019; Miller 2019). However, Sergeeva et al. (2023) found that more than half of individuals still adjusted their decisions based on AI advice, even in the absence of clear, plausible, or convincing explanations. Importantly, current literature on human–AI teaming lacks differentiation between optional or mandatory AI advice and does not explore the relationship between these types of advice and AI convergence.

In summary, although numerous studies examine and contrast the likelihood of advice acceptance based on the source (human vs. AI), context, and individual differences (e.g., Himmelstein and Budescu 2023; Jussupow, Benbasat, and Heinzl 2024), current research on human–AI teaming has largely overlooked how individuals respond to congruent or incongruent signals from human and AI teammates. A notable exception is a study by Xu, Benbasat, and Cenfetelli (2020)



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that investigates how the convergence of recommendations from various sources (AI, human experts, and nonexperts) influences AI convergence, finding that experts recommendations that were congruent with AI were more likely to lead to AI convergence than nonexpert recommendations. Ultimately, this research contributes to this growing literature by investigating how AI signals heuristics (observability and fit) and the (in)congruency of signals across human and AI teammates influence decision-making and the likelihood of AI convergence in different contexts.

## 3 | Theoretical Background and Hypotheses

### 3.1 | Signaling Theory and AI Convergence in Human–AI Teams

To investigate AI convergence in organizational settings, we apply signaling theory (Connelly et al. 2011; Drover, Wood, and Corbett 2018). Signaling theory has demonstrated a significant impact across various areas of organizational studies, including strategic management (Bergh et al. 2014; Park and Patel 2015; Plummer, Allison, and Connelly 2016), human resource management (Guest et al. 2021; Wang, Zhang, and Wan 2021), entrepreneurship (Allison, McKenny, and Short 2013), leadership (Appels 2022; Matthews et al. 2022), diversity initiatives (Leslie 2019), and crowdfunding (Kleinert and Mochkabadi 2022; Steigenberger and Wilhelm 2018). Signaling theory emerged from economics, finance, and cognitive psychology and is useful for investigating behaviors in an environment characterized by uncertainty and information asymmetries (Adam et al. 2022). Unlike most studies utilizing signaling theory to investigate signals as means of external communication, we apply signaling theory to the team context to explain how team members' attributes and actions transmit signals within a team and how those signals can establish credibility and consensus between human and AI team members.

AI systems can emulate complex human behavior, reasoning, and learning (Russell and Norvig 2003). To make decisions in human–AI teams (Bankins et al. 2024), information is distributed within the team through individual signals and signal sets (Drover, Wood, and Corbett 2018). Signaling is especially important in human–AI teaming because AI does not have agency, mutual trust, or emotions that traditional forms of team coordination rely on (Mathieu et al. 2017; Troth et al. 2012). Therefore, the signals sent by AI team members are the primary means by which AI contributes in human–AI teams. Beyond the potential impact of AI signals, we argue that our understanding of the mechanisms governing how employees' decisions align with signals from both human and AI sources in human–AI teams remains incomplete, necessitating a nuanced application of signaling theory.

In human–AI teams, convergence between human and AI team members on critical decisions can manifest in various forms, and following the suggestion of AI is not a given (Marti, Lawrence, and Steele 2024). Three distinct patterns emerge in human–AI teams: (1) human teammates may align and converge with an AI teammate on a specific signal, decision, or evidence; (2) human teammates may diverge with an AI teammate; and (3) in teams involving multiple individuals and an AI, there may be a mixture of convergence and divergence with AI across human teammates. We draw on the cognitive processing approach to signals established by Drover, Wood, and Corbett (2018) to investigate how signal heuristics, signal sets, and volitional action influence the patterns of AI convergence within teams.

To begin, the underlying assumptions of signaling theory are critical to understanding when and how AI convergence in human–AI teams is possible. Signaling theory's impact comes from its ability to explain what does (and what does not) constitute an effective signal. Likewise, signaling theory identifies the situations where a signaling perspective holds the most value. Ultimately, effective signals reduce the uncertainty that exists when two or more entities have different information, a fair assumption when considering the nature of predictive and generative AI.

At a minimum, for signals to be effective in teams, three things must happen. First, signals must catch the attention of team members or be seen and processed at some minimum threshold. Second, signals must provide useful, relevant information that the team needs. Third, the situation must include a level of uncertainty or missing information that a signal can help clarify (Connelly et al. 2011; Drover, Wood, and Corbett 2018). In other words, in human–AI teams, when completing tasks that cannot be accomplished with perfect certainty, individuals have two paths they can walk. First, they can take a risk and charge forward with a decision without clarity or certainty. Second, they can seek out and process relevant signals in their environment that make their task clearer and reduce uncertainty before they make a decision.

Logistically, signaling in human–AI teams contains three core elements: sender(s) (team members, be they human or AI conveying the results of their analysis and prescribing a course of action), receiver(s) (team members with ultimate authority to take an action), and signals (observable cues that convey useful information to decision-makers) (Spence 1973). Previous research suggests that there are two main characteristics of individual signals: observability and fit, which influence their impact (Connelly et al. 2011). Signal sets can be congruent or incongruent depending on whether signals convey the same information or not, and they combine to create the collective positive or negative valence of a signal set.

In human–AI teams, signal observability refers to the extent to which team members notice a signal (Connelly et al. 2011). If too many signals are sent in a team at once, or a signal is not readily observed by team members, then it is unlikely that a signal will have any impact. Logically, it follows that if an AI signal is unobservable, goes unnoticed, or is ignored, AI convergence would be independent of the AI team member and will trend toward chance or unrelated factors.

Signal fit describes the extent to which a signal can truthfully minimize information asymmetries and reduce uncertainty for a specific team member (Bafera and Kleinert 2022; Vanacker et al. 2020). This concept is important in understanding situations



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

when a signal may not actually be an accurate representation of underlying value such as in an unreliable signal or dishonest signal (Busenitz, Fiet, and Moesel 2005; Cohen and Dean 2005; Davila, Foster, and Gupta 2003). Historically, the literature has used various terms to identify the extent to which a signal fits, is credible, and corresponds to quality, including the terms signal strength, intensity, and clarity (for a review, see Connelly et al. 2011).

Two elements of human–AI teaming are important to recognize when it comes to signal fit. First, bias in training data for algorithms can create variability in signal fit depending on the context of the task (Bolukbasi et al. 2016; Buolamwini and Gebru 2018; Çalışkan, Bryson, and Narayanan 2017; Grother, Ngan, and Hanaoka 2019). Therefore, AI signals may be more or less credible depending on the task being solved, and this may be known or unknown by human team members. Second, team members may calibrate signals differently, giving a signal with different perceived fit or meaning within a team (Branzei et al. 2004).

Whereas observability is about a human team member noticing signals from an AI team member, signal fit is largely about how AI signals are interpreted and translated into meaning. Previous research theoretically distinguishes signals with strong and weak fit (Connelly et al. 2011) and proposes that strong signals (those with strong fit) are more likely to evoke a receiver's desired judgmental confidence (a threshold that must be crossed to make a decision) through heuristic processing as human receivers go through an automatic cognitive process of signal interpretation (Drover, Wood, and Corbett 2018). A strongly fitting signal sent by an AI teammate would be known to be reliable and aligned with relevant information from other sources. For example, if an AI were to suggest a performance rating of an employee that is consistent with their previous years' score and the score provided by their immediate supervisor, the confirmation of that score would move swiftly. In other words, signals with strong fit, and no information to the contrary, will largely be noticed and processed automatically when they are present (Drover, Wood, and Corbett 2018).

In contrast, signals with weak fit make it hard for receivers to automatically interpret information and reach the desired judgmental confidence threshold, and therefore, systematic processing is necessary (Drover, Wood, and Corbett 2018). A signal with weak fit sent by an AI teammate is one that contradicts some other relevant piece of information as would be the case if the algorithm above suggested a performance score that was dramatically different than that suggested by the supervisor. If a signal's fit is too weak, receivers may deem its interpretation and the attainment of the desired judgmental confidence threshold as impossible, abandoning any type of cognitive information processing and interrupting effective decision-making.

Applying these concepts to human–AI teams, strong signal fit and credibility from an AI team member should trigger heuristic processing and thus increase the chance of AI convergence in the form of accepting the AI advice. This reasoning leads to the following hypothesis:

**Hypothesis 1.** *In human–AI teams, AI convergence is more likely when a signal sent by an AI has strong signal fit versus when the signal has weak signal fit.*

### 3.2 | Optional AI and AI Convergence

Previous research on human–AI teaming in organizational settings suggests that the distribution of agency is a pivotal factor that significantly impacts AI convergence in the workplace (Agrawal, Gans, and Goldfarb 2019; Chugunova and Sele 2022; Gagné et al. 2022; Grundke 2024). In particular, humans seem to be particularly averse to AI tools that make decisions independently and do not allow human control, especially in contexts with moral implications and high levels of uncertainty (Dietvorst, Simmons, and Massey 2015; Dietvorst and Bharti 2020; Jussupow, Benbasat, and Heinzl 2020; Longoni, Bonezzi, and Morewedge 2019). However, embedding human autonomy and control into human–AI collaborations may mitigate algorithmic aversion and even make people appreciative of AI team members' recommendations (Bigman and Gray 2018; Chugunova and Sele 2022; Gagné et al. 2022; Logg, Minson, and Moore 2019).

This logic was reinforced in a recent review by Gagné et al. (2022) that outlined how mandatory AI, especially without careful consideration of employee needs and interpretation, can hinder the effective implementation of AI technology in the workplace. As teams and work are designed to incorporate new technology, Gagné et al. (2022) suggest that the employee needs for autonomy, competence, and relatedness (Deci and Ryan 1985) will determine how motivated human team members are to converge with AI team members. Empirical evidence suggests that this may be related to the phenomenon of algorithm appreciation (i.e., when humans are motivated to pay attention to AI advice, make use of it, and may even prefer it to human advice). When a task is perceived as objective or analytical, individuals feel more satisfied converging with AI (Chugunova and Sele 2022).

Ultimately, AI convergence is only possible when human team members notice a signal from an AI team member (Connelly et al. 2011). Combining the self-determination theory (Deci and Ryan 1985) with signaling theory (Drover, Wood, and Corbett 2018), we argue that the option to choose whether to see a signal from an AI teammate will help satisfy an individual's need for autonomy (the sense of being agents of their own behavior rather than mere pawns of external pressures, as discussed by Gagné et al. 2022), and therefore increase the likelihood of AI convergence. However, this prediction is reliant not only on employees having autonomy but exercising that autonomy to collaborate with their AI team members. Self-determination theory also posits that beyond the need for autonomy, which can be fulfilled by the ability to choose to see AI advice, employees also need to satisfy a need for competence—a feeling of effectiveness and mastery over their environment (Deci and Ryan 1985; Gagné et al. 2022). When an employee faces uncertainty and chooses to see signals from their AI teammate to enhance their mastery of the environment, their need for competence will be satisfied by the AI advice, thereby making AI convergence more likely. In contrast,



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

when they choose to ignore or eschew their AI teammate, AI convergence will be mitigated.

In summary, we suggest that when employees choose to see AI advice (optional AI advice), rather than being forced to see it (mandatory AI advice), an AI signal takes on two important qualities. First, employees have chosen to attend to a cue in their environment and have exerted agency to satisfy their need for autonomy. Second, the signal becomes costlier as employees now have made a choice to observe and attend to a signal they could have ignored, and their need for competence becomes satisfied from processing the extra information. Taken together, we argue that in human–AI teams, optional AI advice, when observed, should increase AI convergence. This reasoning leads to the following hypothesis:

**Hypothesis 2.** *In human–AI teams, AI convergence is more likely when observing the AI signal is optional versus when observing the AI signal is mandatory, given human team members choose to see AI advice.*

### 3.3 | Signal Sets and AI Convergence

To this point, we have considered the processes of becoming attentive to individual signals that are either optional or have strong or weak fit. Interestingly, signal receivers may not always interpret a set of signals in line with the message intended by the sender, especially when experiencing high levels of uncertainty (Bergh et al. 2014). As signals become more complex, various signals and their components can become bundles that are interpreted together as sets. Such signal sets produce signal set valence, which is defined as "the collective volume and strength of the selected organizational signals and their assigned weight" (Drover, Wood, and Corbett 2018, 218). In other words, when facing information asymmetry and uncertainty, employees will have to make inferences from multiple cues or embedded elements that have different pieces of information (Drover, Wood, and Corbett 2018). Like signals, signal sets can have an array of qualities that make them more or less effective in reducing uncertainty and improving employee confidence in their decisions. There are three possible configurations of signal set congruence: (1) uniform congruence (represents a signal set with either positive or negative signals, depicting a uniform and consistent positive or negative valence across components of the set); (2) imbalanced incongruence (a signal set consists of both positive and negative signals, and the signal set valence is positively or negatively skewed); and (3) balanced incongruence (a signal set comprises both positive and negative signals with equal competing valences) (Drover, Wood, and Corbett 2018).

In human–AI teams, a signal set includes the prescriptions forwarded by the AI and the prescriptions forwarded by a human team member. When these components form a signal set that is congruent (i.e., where the human and the AI team members both signal the same decision), there will be enhanced AI convergence. When signals create a signal set that is incongruent (i.e., where the human and the AI team members both signal different decisions), AI convergence will be less likely. As repeated signals have been shown to increase signal impact (Connelly et al. 2011), we believe this effect will hold regardless of whether signals have strong fit or whether observing AI signals is chosen and not mandatory. Based on these arguments, we hypothesize the following:

**Hypothesis 3.** *In human–AI teams, AI convergence is more likely when the signal set created by human and AI team members is congruent versus incongruent.*

### 3.4 | Research Accountability

Following best research practices, we preregistered Hypotheses 1 and 2 before any data collection started. Hypothesis 3 emerged in response to excellent reviewer feedback and was not preregistered. Findings are replicated across several samples (four studies total) in fully independent samples from employees and leaders working in multiple countries, industries, and roles. To ensure the validity of our experiments, participants from each study were excluded from participating in the following studies using Prolific ID exclusion lists. Research for this paper was approved by the IRB of one of the author's institutions. In addition to this, we have provided all data and code used to produce the results. This includes a transparent data cleaning process where we upload the full dataset with two key cleaning variables. First, a dummy variable identifies which rows and participants are to be dropped from the analysis. Second, we include a column that identifies the reason each row was dropped. Hypotheses were tested using the "melogit" command in Stata 16 with random intercepts based on participant ID. Data and code can be found on the OSF repository: https://osf.io/zjwta/?view_only=74a15f71b21c4b0abb69f3e5de543537 and our preregistration can be found here: https://osf.io/yt5mc/?view_only=f5efa8fc9d2041f4a44dd4b6fc7f6bfe.

## 4 | Research Methods

### 4.1 | Study 1 Method

#### 4.1.1 | Participants

We recruited an initial sample of 304 employees from the United Kingdom and the United States to take part in our study using the Prolific platform. Participants were recruited until 300 had completed the experiment without failing attention checks. Participants had to be at least 18 years of age, be employed part-time or full-time, and have a good command of English. A total of 304 participants followed the link to participate in the study. Prior to participants proceeding to the main study, they were provided a tutorial specific to the condition they were randomly assigned to. Participants were able to practice the task with two sample image pairings. We removed four participants who failed to provide consent or failed process checks by not recalling basic information after the tutorial. This resulted in 300 individuals who were allowed to proceed to the full experiment and were paid 2.90 Pounds in exchange for their time. During the experiment, an additional 20 participants failed attention checks and were omitted from the analysis. This sample size surpassed the power requirements to detect a medium effect ($\eta_p^2 = 0.25$; Cohen 1988) with 80% power. To ensure the generalizability of our findings and adhere to equitable research practices



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Offenwanger et al. 2021), we aimed for a balanced sample in terms of gender and race because of the nature of the task, which involved matching faces of Black and White men and women. The participants were on average 42.835 years old (*SD* 14.37). Approximately 51% of our participants identified as female and 49% of participants identified as male. Approximately 50.36% of participants identified as Black and 48.64% identified as White.

#### 4.1.2 | Procedure

Each participant was randomly assigned into one of four experimental conditions following a 2 (signal fit: strong versus weak) × 2 (AI advice: optional versus mandatory) between-subject design. Within each condition, participants completed a series of face-matching tasks in an employee identification scenario. Participants were instructed to play the role of a security officer who needs to identify if a person attempting to enter a highly secure facility is indeed a current employee with clearance to enter the building. The experimental setup was a two-step process. In the first step, participants were asked to review the image of the person seeking entry and compare this image to the most similar image in the employee image database deciding to grant or deny entry. In the second step, depending on their condition, participants were sent a signal with strong or weak fit and a signal of optional or mandatory AI advice. Participants had to decide whether the two images were a match (i.e., grant entry) or not a match (i.e., deny entry). In total, participants evaluated a series of 24 pairs of pretested images of human faces that were evenly balanced between races (Black vs. White) and gender (female vs. male). After evaluating each pair of images, participants reported their level of confidence. After finishing the facial recognition task, participants went through a series of demographic questions and additional measures (e.g., explicit gender bias).

The pairs of facial images employed in Study 1 were designed using the color version of the Facial Recognition Technology (FERET) database (Phillips et al. 1998). To select pairs of images for the experiment, the photos were initially evaluated by a face application programming interface (API) that used an algorithm to produce a similarity score between 0 (not similar at all) and 1 (very similar). However, the API had difficulties in creating viable pairs of images for Black men and women in comparison with White men and women. Therefore, images for Black men and women were selected manually from the database rather than using the algorithmically generated scores. To ensure a similar level of difficulty for the selected images, we conducted a preliminary study. We recruited 160 participants on Prolific balanced by gender (men/women) and race (Black/White). Participants were asked to determine if the pairs of Black/White men and women images were a “match” or “no match” and to rank their confidence after each pair of images. No other information was provided. Based on the difficulty of the images (e.g., how many times the pairs of images were matched or not matched correctly) and the confidence ratings, we chose 24 images (six images for White men, six images for White women, six images for Black men, and six images for Black women) of equal difficulty to be used in the main experiment. Results of this analysis can be provided upon request.

**4.1.2.1 | Manipulation of Signal Fit.** In the strong signal fit (coded as 0), participants were presented with a pair of images next to an employee badge as well as the decision made by the AI tool (facial recognition system), shown as *match* or *no match* (see Figure 1). In the weak signal fit (coded as 1), participants

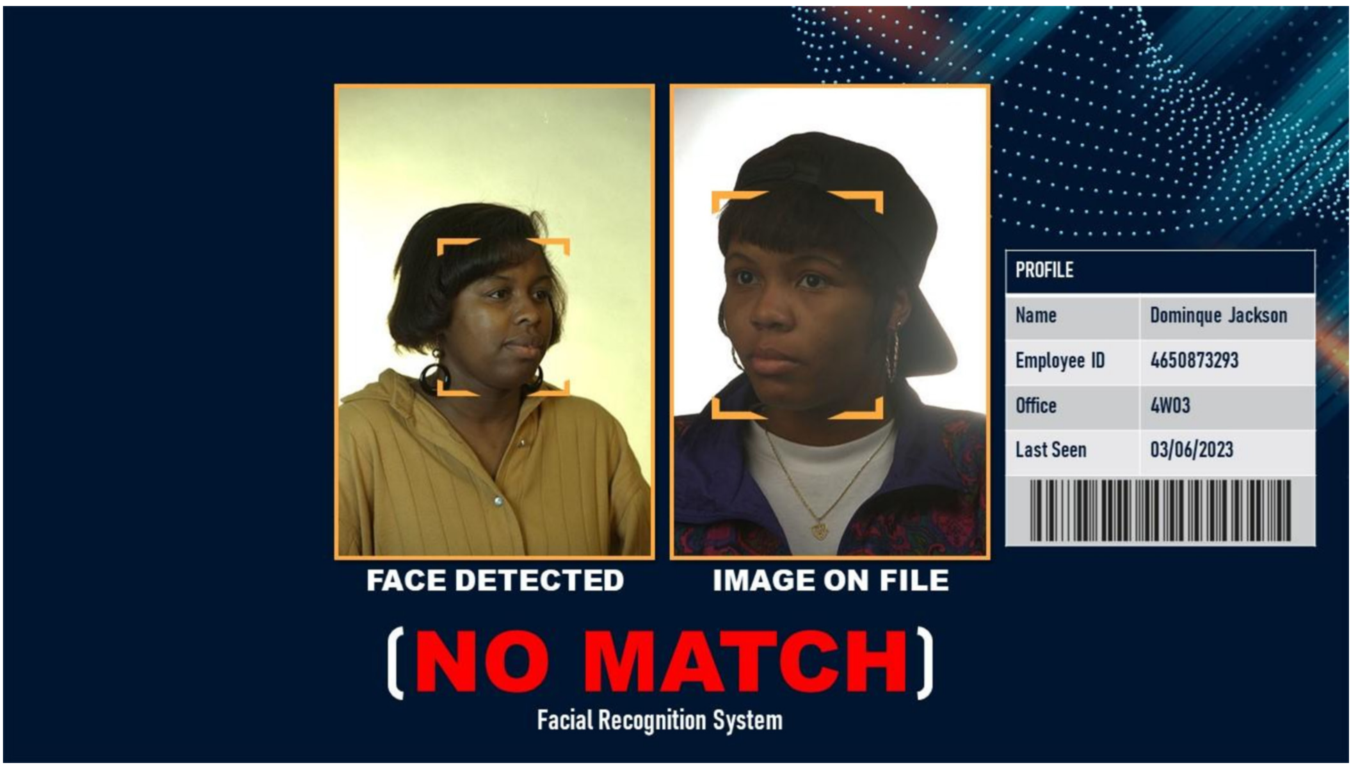


**FIGURE 1** | User interface showing strong signal fit condition in Studies 1 and 2.



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

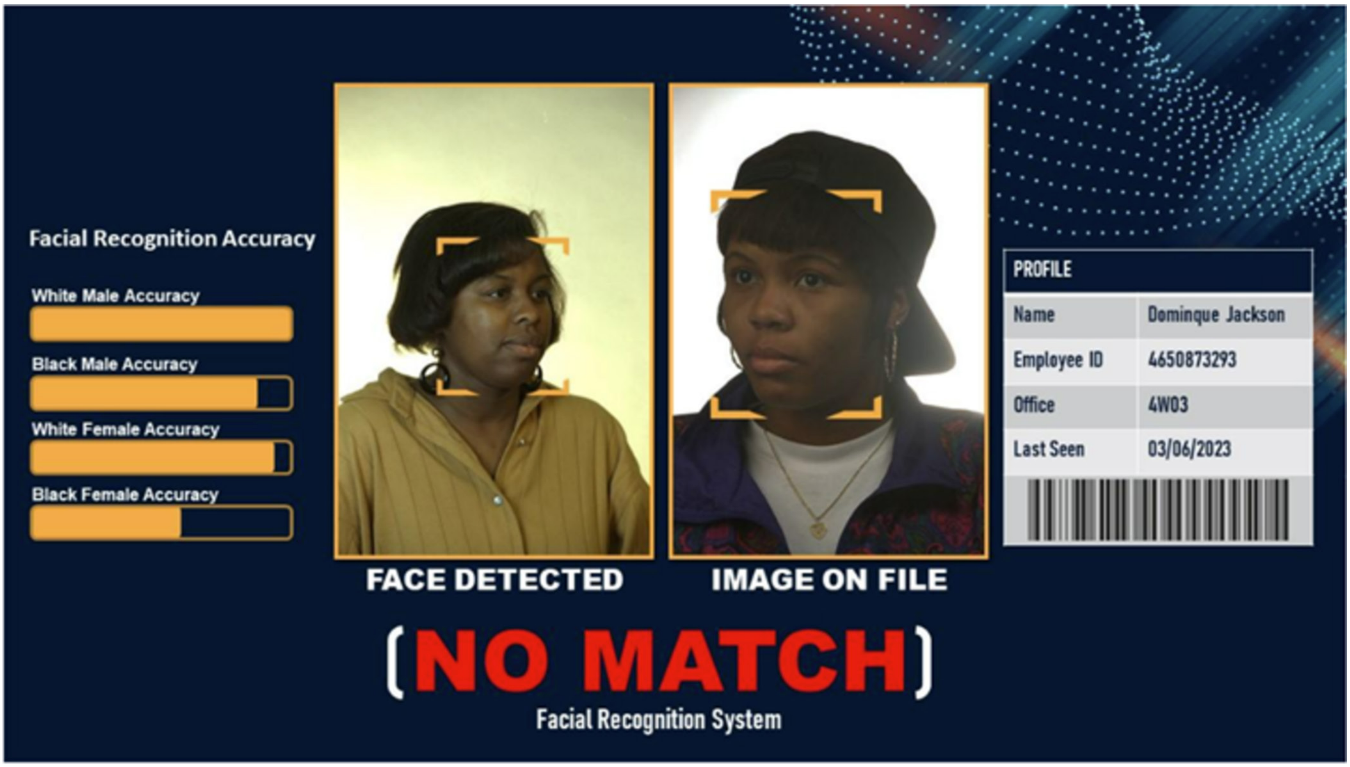


**FIGURE 2** | User interface showing weak signal fit condition in Studies 1 and 2.

were shown the AI decision (match or no match) plus additional information about the level of accuracy of the AI for different demographics (White Male/Black Male/White Female/Black Female) represented as bar graphs (see Figure 2). This information weakened the signal by limiting the credibility of AI advice. Visual representation for the AI accuracy for different demographics was derived from previous research on gender and racial disparity from facial recognition system vendors (Buolamwini and Gebru 2018; Grother, Ngan, and Hanaoka 2019). $\chi^2$ tests revealed that participants accurately recalled whether they were provided with AI accuracy ($\chi^2=100.116$, $p=0.000$). Logistic regressions also showed that only the variable capturing the weak fit signal condition was a significant predictor of recalling that participants had information about AI accuracy ($p=0.000$). There was no effect of the optional AI advice ($p=0.640$) condition or the interaction between optional AI advice and signal fit ($p=0.653$).

#### 4.1.2.2 | Manipulation of Optional AI Advice.

In the optional AI advice condition (coded as 1), participants were asked if they would like additional assistance from the AI to help them make their decision. If participants opted to receive AI advice, they were prompted with information depending on the condition they were in (strong signal fit–AI match/no match vs. weak signal fit–AI match/no match plus AI accuracy for different demographics). If participants did not opt for AI advice, they received no information from the facial recognition system and made an unaided decision for that image pair. In the mandatory AI condition (coded as 0), participants were not asked if they would like additional assistance. Instead, they were automatically presented with either a strong fit or weak fit signal, depending on their condition. $\chi^2$ tests revealed that participants accurately recalled whether they had the option to solicit AI advice ($\chi^2=76.786$, $p=0.000$). Logistic regressions also showed that only the variable capturing the optional AI advice condition was a significant predictor of recalling that participants had the option to solicit AI advice ($p=0.000$). There was no effect of the signal fit ($p=0.413$) condition or the interaction between optional AI advice and signal fit ($p=0.778$).

### 4.1.3 | Measures

#### 4.1.3.1 | Grant/Deny Access.

For each image pair, participants were asked to either grant or deny access if they believed the pairs of images to be a match (or mismatch), respectively. Each match/grant decision was captured using a dichotomous variable where 1 = match/grant access and 0 = mismatch/deny access.

#### 4.1.3.2 | Confidence in Decision.

Following each decision, participants reported their level of confidence in their decision using a sliding scale ranging from 0 to 100, with 0 being not confident at all to 100 being very confident in decision.

#### 4.1.3.3 | Decline to Solicit AI Advice.

For participants in the optional AI advice condition, a dichotomous variable captured whether or not participants chose to obtain additional assistance from the facial recognition system. This variable was coded as a 0 if they selected "YES—I would like to solicit the facial recognition system as I am not very confident in my current decision" and 1 if they selected "NO—I am confident in my decision and do not require further assistance."

#### 4.1.3.4 | AI Convergence.

A dichotomous variable captured whether or not each participant's match/grant access



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

decision was the same as the advice provided by the AI. A value of 1 indicated that the employee converged with the AI, making either the same grant access decision (when the pairs were indeed a photo of the same person) or the same deny access decision (when the photo pairs included images of different individuals). When participants diverged from the AI, the value of this dichotomous variable was set as 0.

**4.1.3.5 | AI Knowledge.** To evaluate participants' knowledge of AI, we modified a measure developed by Maier, Jussupow, and Heinzl (2019). Participants were asked to report their level of understanding, level of interaction, and level of interest in regard to facial recognition systems using a slider from 0 ("I've never heard of facial recognition systems") to 100 ("I've heard a lot about facial recognition systems").

**4.1.3.6 | Trust in AI.** Trust in AI was measured using a scale developed by Merritt (2011) and modified to focus on facial recognition systems. Sample items include, "I believe facial recognition systems are competent performers," "I have confidence in the advice given by facial recognition systems," and "I can rely on facial recognition systems to do its best every time I take its advice." Items range from 1 (strongly disagree) to 5 (strongly agree) ($\alpha = 0.87$).

**4.1.3.7 | Explicit Gender Bias.** Explicit gender bias was measured using a scale developed by Swim et al. (1995). Sample statements include "On average, people in our society treat men and women equally" and "Society has reached the point where women and men have equal opportunities for achievement." Items range on a Likert scale from 1 (strongly disagree) to 7 (strongly agree) ($\alpha = 0.85$).

**4.1.3.8 | Explicit Racial Bias.** Explicit racial bias was measured using a scale developed by Uhlmann, Brescoll, and Machery (2010). Sample statements include "If your personal safety is at stake, it's sensible to avoid members of ethnic groups known to behave more aggressively" and "Law enforcement officers should act as if members of all racial groups are equally likely to commit crimes." Items range on a Likert scale from 1 (strongly disagree) to 7 (strongly agree) ($\alpha = 0.76$).

**4.1.3.9 | Demographics.** We collected data on age, gender, race, level of education, employment status, earned income, work sector, country of residency, community (e.g., urban, suburban, and rural), and level of English proficiency.

## 4.2 | Study 1 Results

Table 1 displays the means, standard deviations, and descriptive statistics of our study variables. Hypothesis 1 suggested that a strong signal fit versus a weak signal fit from AI would increase AI convergence in a high-stakes scenario. Hypothesis 2 suggested that having the option to get AI advice would likewise increase AI convergence, given that a human chooses to see the optional AI advice. We tested these hypotheses using a mixed effects logistic regression command "melogit" in Stata 16

**TABLE 1** | Study 1 means, correlations, and standard deviations.

| | Variable | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | AI convergence | 0.83 | 0.38 | 1.00 | | | | | | | | | | |
| 2 | Weak signal fit | 0.50 | 0.50 | −0.03 | 1.00 | | | | | | | | | |
| 3 | Optional AI | 0.52 | 0.50 | −0.06 | 0.05 | 1.00 | | | | | | | | |
| 4 | Decline to solicit | 0.16 | 0.37 | −0.28 | 0.08 | 0.42 | 1.00 | | | | | | | |
| 5 | Female | 0.51 | 0.50 | −0.01 | −0.11 | 0.01 | 0.04 | 1.00 | | | | | | |
| 6 | Black participant | 0.50 | 0.50 | −0.03 | 0.01 | 0.01 | 0.11 | 0.03 | 1.00 | | | | | |
| 7 | AI knowledge | 55.58 | 20.93 | 0.00 | −0.12 | −0.03 | 0.07 | 0.02 | 0.30 | 1.00 | | | | |
| 8 | Trusting disposition | 4.66 | 1.42 | 0.04 | −0.16 | −0.06 | −0.06 | 0.03 | −0.11 | 0.11 | 1.00 | | | |
| 9 | Human–computer trust | 3.13 | 0.83 | 0.06 | −0.05 | 0.02 | 0.05 | −0.07 | 0.10 | 0.38 | 0.18 | 1.00 | | |
| 10 | Explicit racism | 2.70 | 1.22 | −0.01 | −0.02 | −0.10 | −0.03 | 0.00 | −0.03 | 0.11 | −0.12 | 0.20 | 1.00 | |
| 11 | Explicit sexism | 3.33 | 1.26 | −0.02 | −0.05 | −0.12 | −0.01 | −0.02 | −0.07 | 0.15 | 0.09 | 0.26 | 0.34 | 1.00 |

*Note:* *N* Level 2 = 280; *N* Level 1 = 6720. All correlations above 0.025 in absolute value are significant at $p < 0.05$.



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Statacorp 2019) with random intercepts for each participant. The nature of the data, which involve repeated binary decisions nested within participants, required the use of multilevel logistic regression to address issues of nonindependence and our binary dependent variable (see Guo and Zhao 2000 for a review). We first examined the relationships between the main effects of our experimental conditions, comparing the AI convergence of participants who had received a strong versus weak signal fit, as well as those who could or could not choose to solicit optional AI advice. Next, we added a dummy variable capturing whether participants chose to see AI advice.

For the signal fit, results show that there was no significant effect of a strong signal fit on the likelihood of AI convergence (*Odds ratio* = 0.797, $p$ = 0.161). For optional AI advice, results show a significant negative effect on AI convergence when AI was optional versus when it was mandatory (*Odds ratio* = 0.630, $p$ = 0.005). Next, we added an additional covariate to our model, which captured whether or not participants who had the option to solicit AI advice actually chose not to see it. Results show that when accounting for whether or not a participant actually chose to see optional AI advice, the effect of signal fit remained unchanged (*Odds ratio* = 0.884, $p$ = 0.413). Thus, Hypothesis 1 was not supported as the AI signal fit did not impact AI convergence.

In contrast, when accounting for whether a participant actually chose to see optional AI advice, the effect of having the option to solicit AI advice became positive, increasing the odds of convergence (*Odds ratio* = 1.508, $p$ = 0.011), whereas the effect of choosing not to see the AI advice had a large negative effect, decreasing the odds of making the same decision as the AI (*Odds ratio* = 0.126, $p$ = 0.000). Thus, Hypothesis 2 was partially supported with the caveat that the participants' choice to actually solicit AI advice was a critical factor in determining the impact of optional AI. Importantly, for interpreting our results in light of Hypothesis 2, controlling for whether or not one exercises the choice to see AI advice leads to the interpretation that ceteris paribus, having the option does have a strong, positive impact on taking AI advice but this is contingent on the participant seeing the AI advice.[1]

#### 4.2.1 | Supplementary Analyses

We tested the interaction between the two main effects. We found no interaction between signal fit and optional AI advice on AI convergence by human teammates (signal fit main effect: *Odds ratio* = 0.832, $p$ = 0.404; optional AI advice main effect: *Odds ratio* = 1.422, $p$ = 0.114; interaction: *Odds ratio* = 1.120, $p$ = 0.706). Given that the results of our study were unexpected, we conducted a series of abductive robustness checks and supplementary analyses to better understand the nature of the data and of our experiment. Critically, we measured participants' confidence in each decision they made as this has been shown to be a key driver in AI-supported decision-making (Chong et al. 2022). Although not originally preregistered, any confidence level less than 100% would suggest participants experience uncertainty, which is essential under signaling theory. Therefore, we controlled for participants' level of confidence and other variables that would lend support for our theorizing. In these analyses, we explored whether confidence, explicit racial bias, explicit gender bias, participant gender, participant race, and trust in AI (a component of algorithmic appreciation) altered the effects of our experiment. Results of this analysis show no qualitative difference in our pattern of results when adding these controls; however, participant's racial bias had a negative impact on AI convergence, whereas confidence and trust in AI significantly improved the chance that employees took AI team member's advice (see Tables 2 and 3).

### 4.3 | Study 1 Discussion

The results of Study 1 were quite surprising. Hypothesis 1 was not supported suggesting that signal fit did not impact human employees' tendency to converge with AI in their decisions. On its own, signal fit does not appear to play a major role in the AI convergence in the workplace, at least not in a dyadic partnership in the facial recognition system context. That said, the results supported Hypothesis 2 suggesting that an optional AI signal is crucial for the effective human–AI teaming in organizational settings. The effect of optional AI became positive when controlling for whether receivers choose to see AI advice or not. This is an important finding in regard to the design of

**TABLE 2** | Study 1 results.

| Variable | Model 1 main effects only: Odds ratio | SE | Model 2 main effects plus solicit decision: Odds ratio | SE | Model 3 interaction plus solicit decision: Odds ratio | SE |
|---|---|---|---|---|---|---|
| Weak signal fit | 0.80 | 0.13 | 0.88 | 0.13 | 0.83 | 0.18 |
| Optional AI advice | 0.63** | 0.1 | 1.51* | 0.24 | 1.42 | 0.32 |
| Decline to solicit | | | 0.13** | 0.01 | 0.13** | 0.01 |
| Weak signal fit × Optional AI advice | | | | | 1.12 | 0.34 |
| Intercept | 9.99** | 1.49 | 9.01** | 1.24 | 9.28** | 1.48 |
| var (participant ID) | 1.37 | 0.19 | 1.10 | 0.16 | 1.10 | 0.16 |

*Note:* $N$ Level 2 (Participants) = 280; $N$ Level 1 (Observations) = 6720.
**$p < 0.01$.
*$p < 0.05$.
†$p < 0.1$.



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 3** | Study 1 supplementary analysis results.

| Variable | Model 1 main effects only | | Model 2 main effects plus solicit decision | | Model 3 interaction plus solicit decision | |
|---|---|---|---|---|---|---|
| | Odds ratio | SE | Odds ratio | SE | Odds ratio | SE |
| Weak signal fit | 0.78 | 0.13 | 0.90 | 0.13 | 0.88 | 0.20 |
| Optional AI advice | 0.55** | 0.09 | 1.48* | 0.24 | 1.46† | 0.34 |
| Decline to solicit | | | 0.09** | 0.01 | 0.09** | 0.01 |
| Weak signal fit × Optional AI advice | | | | | 1.03 | 0.31 |
| Confidence in decision | 01.01** | 0.00 | 1.03** | 0.00 | 1.03** | 0.00 |
| Female participant | 0.89 | 0.16 | 0.96 | 0.15 | 0.96 | 0.14 |
| Black participant | 0.72† | 0.12 | 0.82 | 0.13 | 0.82 | 0.13 |
| Explicit racial bias | 0.94 | 0.07 | 0.94 | 0.06 | 0.94 | 0.06 |
| Explicit gender bias | 0.91 | 0.06 | 0.93 | 0.06 | 0.93 | 0.06 |
| AI knowledge | 0.99 | 0.00 | 1.00 | 0.00 | 0.99 | 0.00 |
| Trust in AI system | 1.40** | 0.15 | 1.35** | 0.14 | 1.33** | 0.14 |
| Intercept | 2.00 | 0.86 | 0.81 | 0.32 | 0.82 | 0.35 |
| var (participant ID) | 1.29 | 0.18 | 1.02 | 0.15 | 1.02 | 0.15 |

*Note:* $N$ Level 2 (Participants) = 280; $N$ Level 1 (Observations) = 6720.
** $p < 0.01$.
* $p < 0.05$.
† $p < 0.1$.

human–AI teams that reinforces the "augmentation thesis," while favoring a collaborative model of hybrid decision-making. Ultimately, these data suggest that we may need to reject the oppositional framing of AI versus human decision-making to optimize effective human–AI teams. Finally, the effect of optional AI was independent of the signal fit about the AI partner's signal that was sent to receivers.

The results of Study 1 demonstrated clearly that we needed to replicate our findings. It is possible that the data represented a one-time finding that was idiosyncratic to the specific data set. Beyond this, we determined that it was necessary to address a potential confound in our experimental design. Namely, in Study 1, we asked participants in the optional AI condition to solicit optional AI advice only if they were not confident in their initial decision (see Solicit AI Advice measure). Although this design is aligned with the stage of attending to observable signals as well as with the stage of signal interpretation described by signaling theory, it nevertheless could be the case that we introduced demand effects, or otherwise altered the behavior of participants by making confidence directly tied to the decision to see AI advice in the optional AI conditions.

## 4.4 | Study 2 Method

### 4.4.1 | Participants and Procedure

We followed the same procedures for Study 2 as in Study 1 with two exceptions. For Study 2, we constrained participant location to be within the United States. We also removed any language about confidence when asking participants in the optional AI conditions if they wanted to see optional AI advice. A total of 296 participants followed the link to participate in the study. We removed seven participants who failed to provide consent or failed to recall basic information after the tutorial (process checks). This resulted in 289 individuals interested who were allowed to proceed to the full experiment and were paid 2.90 pounds in exchange for their time. These participants were on average 44.18 years old (*SD* 13.26). Approximately 48.6% of our participants identified as female, and 51.4% of participants identified as male. Approximately 49% of participants identified as Black and 51% identified as White.

### 4.4.2 | Measures

The same measures that were used in Study 1 were used in Study 2 with one exception. Participants in the optional AI conditions were asked if they wanted to see optional AI advice and given either YES or NO to answer. The internal consistency of our measures was adequate ($\alpha$ Trust in AI = 0.94; $\alpha$ Explicit Gender Bias = 0.91; $\alpha$ Explicit Racial Bias = 0.78).

### 4.4.3 | Manipulation Checks

Like in Study 1, $\chi^2$ tests revealed that participants accurately recalled whether they had the option to solicit AI advice



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

($\chi^2 = 83.164$, $p = 0.000$) and the signal fit ($\chi^2 = 74.176$, $p = 0.000$). Logistic regressions also showed that only the variable capturing each experimental condition was a significant predictor of recalling information about that condition ($p = 0.000$ in each case). When recalling whether there was optional AI advice, there was no effect of signal fit ($p = 0.451$) or the interaction between optional AI advice and signal fit ($p = 0.472$). The parallel check for the signal fit condition followed the same pattern. There was no effect of the optional AI advice ($p = 0.279$) condition or the interaction between optional AI advice and signal fit condition ($p = 0.985$).

### 4.5 | Study 2 Results

Table 4 displays the means, standard deviations, and descriptive statistics of our Study 2 variables. For the second time, results show that for signal fit, there was no significant effect of additional information that weakened the signal on the likelihood of AI convergence (*Odds ratio* = 0.827, $p = 0.218$). Furthermore, for optional AI advice, results again show a significant, negative effect in the likelihood of AI convergence when AI advice was optional versus when it was mandatory (*Odds ratio* = 0.705, $p = 0.024$). Results again show the same pattern as in Study 1 when accounting for whether or not a participant actually chose to see AI advice. When controlling for a participant's decision to see AI advice when given the choice, the effect of signal fit remained nonsignificant (*Odds ratio* = 0.900, $p = 0.475$), whereas the effect of having the option to see AI advice became positive, increasing the odds of AI convergence (*Odds ratio* = 1.400, $p = 0.034$). Again, the effect of ignoring the option for AI advice had a large negative effect, decreasing the odds that participants made the decision as the AI suggested (*Odds ratio* = 0.159, $p = 0.000$). Thus, Hypothesis 1 was not supported, and Hypothesis 2 was supported for a second time.

#### 4.5.1 | Study 2 Supplementary Analysis

Similar to Study 1, we controlled for participants' level of confidence and other variables that would lend support for our theorizing. Results of this analysis again show no qualitative difference in our tests of Hypotheses 1 and 2 when adding the controls of confidence, trust in AI, racial bias, and gender bias (see Tables 5 and 6). Similar to Study 1, participants' confidence and participants' trust in their AI teammate enhanced the likelihood of AI convergence. Additionally, women participants were more likely to converge with AI advice compared with men. Finally, the interaction between signal fit and optional AI advice was not significant, similar to Study 1 (signal fit main effect: *Odds ratio* = 0.884, $p = 0.568$; optional AI advice main effect: *Odds ratio* = 1.372, $p = 0.142$; interaction: *Odds ratio* = 1.034, $p = 0.909$).

### 4.6 | Study 2 Discussion

The results of Study 2 replicated Hypothesis 2 while removing potential demand effects inherent to the design in Study

**TABLE 4** | Study 2 means, correlations, and standard deviations.

| | Variable | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | AI convergence | 0.83 | 0.38 | 1.00 | | | | | | | | | | |
| 2 | Weak signal fit | 0.49 | 0.50 | −0.02 | 1.00 | | | | | | | | | |
| 3 | Optional AI | 0.53 | 0.50 | −0.04 | −0.02 | 1.00 | | | | | | | | |
| 4 | Decline to solicit | 0.15 | 0.35 | −0.23 | 0.05 | 0.39 | 1.00 | | | | | | | |
| 5 | Female | 0.49 | 0.50 | 0.05 | 0.02 | 0.01 | −0.01 | 1.00 | | | | | | |
| 6 | Black participant | 0.49 | 0.50 | 0.01 | −0.02 | 0.03 | 0.07 | 0.00 | 1.00 | | | | | |
| 7 | AI knowledge | 58.41 | 18.66 | 0.00 | 0.11 | −0.08 | 0.01 | −0.16 | 0.19 | 1.00 | | | | |
| 8 | Trusting disposition | 4.21 | 1.60 | −0.03 | −0.03 | −0.08 | −0.09 | −0.12 | −0.18 | 0.16 | 1.00 | | | |
| 9 | Human–computer trust | 3.09 | 0.87 | 0.08 | 0.09 | −0.04 | 0.02 | −0.13 | 0.13 | 0.32 | 0.21 | 1.00 | | |
| 10 | Explicit racism | 2.81 | 1.29 | −0.01 | 0.12 | −0.10 | 0.05 | −0.16 | −0.01 | 0.03 | −0.07 | 0.21 | 1.00 | |
| 11 | Explicit sexism | 2.99 | 1.43 | 0.02 | 0.05 | 0.05 | 0.07 | −0.26 | −0.05 | 0.02 | 0.03 | 0.30 | 0.50 | 1.00 |

*Note:* *N* Level 2 = 289; *N* Level 1 = 6936. All correlations above 0.025 in absolute value are significant at $p < 0.05$.



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 5** | Study 2 results.

| Variable | Model 1 main effects only | | Model 2 main effects plus solicit decision | | Model 3 interaction plus solicit decision | |
|---|---|---|---|---|---|---|
| | Odds ratio | SE | Odds ratio | SE | Odds ratio | SE |
| Weak signal fit | 0.83 | 0.13 | 0.90 | 0.13 | 0.88 | 0.19 |
| Optional AI advice | 0.70* | 0.11 | 1.40* | 0.22 | 1.37** | 0.3 |
| Decline to solicit | | | 0.16** | 0.02 | 0.16** | 0.02 |
| Weak signal fit × Optional AI advice | | | | | 1.03 | 0.31 |
| Intercept | 9.09** | 1.3 | 8.45** | 1.15 | 8.52** | 1.35 |
| var (participant ID) | 1.27 | 0.17 | 1.11 | 0.15 | 1.11 | 0.15 |

*Note:* $N$ Level 2 (Participants) = 289; $N$ Level 1 (Observations) = 6936.
$^{**}p < 0.01$.
$^{*}p < 0.05$.
$^{\dagger}p < 0.1$.

**TABLE 6** | Study 2 supplementary analysis results.

| Variable | Model 1 main effects only | | Model 2 main effects plus solicit decision | | Model 3 interaction plus solicit decision | |
|---|---|---|---|---|---|---|
| | Odds ratio | SE | Odds ratio | SE | Odds ratio | SE |
| Weak signal fit | 0.78 | 0.12 | 0.83 | 0.12 | 0.78 | 0.17 |
| Optional AI advice | 0.62** | 0.10 | 1.36* | 0.21 | 1.27 | 0.28 |
| Decline to solicit | | | 0.11** | 0.01 | 0.11** | 0.01 |
| Weak signal fit × Optional AI advice | | | | | 1.13 | 0.33 |
| Confidence in decision | 1.02** | 0.00 | 1.03** | 0.00 | 1.02** | 0.00 |
| Female participant | 1.44* | 0.23 | 1.46* | 0.22 | 1.46** | 0.22 |
| Black participant | 0.96 | 0.15 | 1.06 | 0.16 | 1.06 | 0.16 |
| Explicit racial bias | 0.97 | 0.07 | 1.03 | 0.07 | 1.03 | 0.07 |
| Explicit gender bias | 1.01 | 0.06 | 1.02 | 0.06 | 1.02 | 0.06 |
| AI knowledge | 0.99† | 0.00 | 0.99† | 0.00 | 0.99† | 0.00 |
| Trust in AI | 1.41** | 0.12 | 1.41** | 0.13 | 1.41** | 0.13 |
| Intercept | 0.96 | 0.37 | 0.44* | 0.17 | 0.45* | 0.18 |
| var (participant ID) | 1.16 | 0.16 | 1.01 | 0.14 | 1.01 | 0.14 |

*Note:* $N$ Level 2 (Participants) = 289; $N$ Level 1 (Observations) = 6936.
$^{**}p < 0.01$.
$^{*}p < 0.05$.
$^{\dagger}p < 0.1$.

1. Optional AI advice seemed to improve the likelihood of employees' convergence with AI, but only when participants chose to see the optional AI signal. Similar to Study 1, this effect appeared to be independent of signal fit. Thus far, our studies have tested our first two hypotheses in a dyadic human–AI partnership. Given this study design, we cannot speak to the context of AI teams, where there is more than one human employee working in a team that includes AI. Therefore, in Study 3, we extend our study design to include an additional signal sent from a human teammate.

## 4.7 | Study 3 Method

### 4.7.1 | Participants and Procedure

We followed the same procedures for Study 3 as in Study 2 with three exceptions. First, we reduced the number of images to 16 to reduce participant burden. We selected the images that were most difficult in our previous two studies to retain, so that there would be the highest level of uncertainty present to drive signal salience and interpretation (Drover,



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Wood, and Corbett 2018). Second, we included an additional signal sent by a human colleague, Sam Smith, for every image pair. With this addition, we were able to manipulate the congruence of the signal set composed of human and AI team member signals. The study was balanced, so that half of the signal sets were congruent, where the human and the AI both signaled the same decision. In the other half, the signal from the AI was the opposite of the signal from Sam Smith. Finally, we gave the facial recognition system tool the name Iris AI to match the presentation of the colleague Sam Smith in our decision-making task. A total of 319 participants followed the link to participate in the study. We removed 20 participants who failed to provide consent or failed to recall basic information after the tutorial (process checks). This resulted in 299 individuals interested who were allowed to proceed to the full experiment and were paid 2.90 pounds in exchange for their time. These participants were on average 43.26 years old (*SD* 12.82). Approximately 47.8% of our participants identified as female, 50.2% of participants identified as male, 1.7% identified as transgender, and 0.3% identified as nonconforming. Approximately 49.2% of participants identified as Black/Caribbean/African/Afro Latin and 49.8% identified as White/Caucasian/European.

**4.7.1.1 | Signal Set Congruence.** In the congruent signal set condition (coded as 1), participants were presented with a pair of images next to an employee badge as well as the decision made by Iris AI and Sam Smith, where both signals were the same (see Figure 3). In the incongruent signal set conditions (coded as 0), participants were shown Iris AI's decision (match or no match) plus Sam Smith's decision (match or no match) but the signal sent by Sam Smith and Iris AI was the opposite (see Figure 4).

### 4.7.2 | Measures

The same focal measures that were used in Study 2 were used in Study 3 except where detailed below. The internal consistencies of our measures were adequate ($\alpha$ Trust in AI = 0.94; $\alpha$ Explicit Gender Bias = 0.91; $\alpha$ Explicit Racial Bias = 0.78).

**4.7.2.1 | Algorithm Aversion.** In our review of the literature, it was clear the research on algorithmic aversion relies heavily on experimental results (Jussupow, Benbasat, and Heinzl 2024; Mahmud et al. 2022). Although our previous studies captured the trust participants had in AI systems, to our knowledge, a comprehensive measure of the underlying dimensions of algorithm aversion does not exist. To operationalize this construct, we developed a scale to establish construct validity and psychometric soundness. After conducting a review of the literature, the authors first generated a list of 60 content-valid items capturing four dimensions of algorithm aversion through an iterative process of prompt engineering using ChatGPT prompts (Bail 2024; Götz et al. 2023). These dimensions were Fear of Inaccuracy, Lack of Trust, Fear of Loss of Control, and Resistance to Change (15 items for each dimension). Participants were asked to complete all 60 items at the end of the survey. On analyzing these data using exploratory factor analysis, five factors had eigenvalues greater than 1 and were retained for rotation. Factors were rotated using an oblique rotation allowing factors to correlate. Using the cutoff criteria of 0.4 for item loading, after rotation, the fifth factor consisted mostly of cross loaded items that were ultimately dropped from the scale, resulting in a final scale of 45 items across four factors. Table 7 shows the items and factor loadings for our new measure. The internal consistency of this measure and its subdimensions

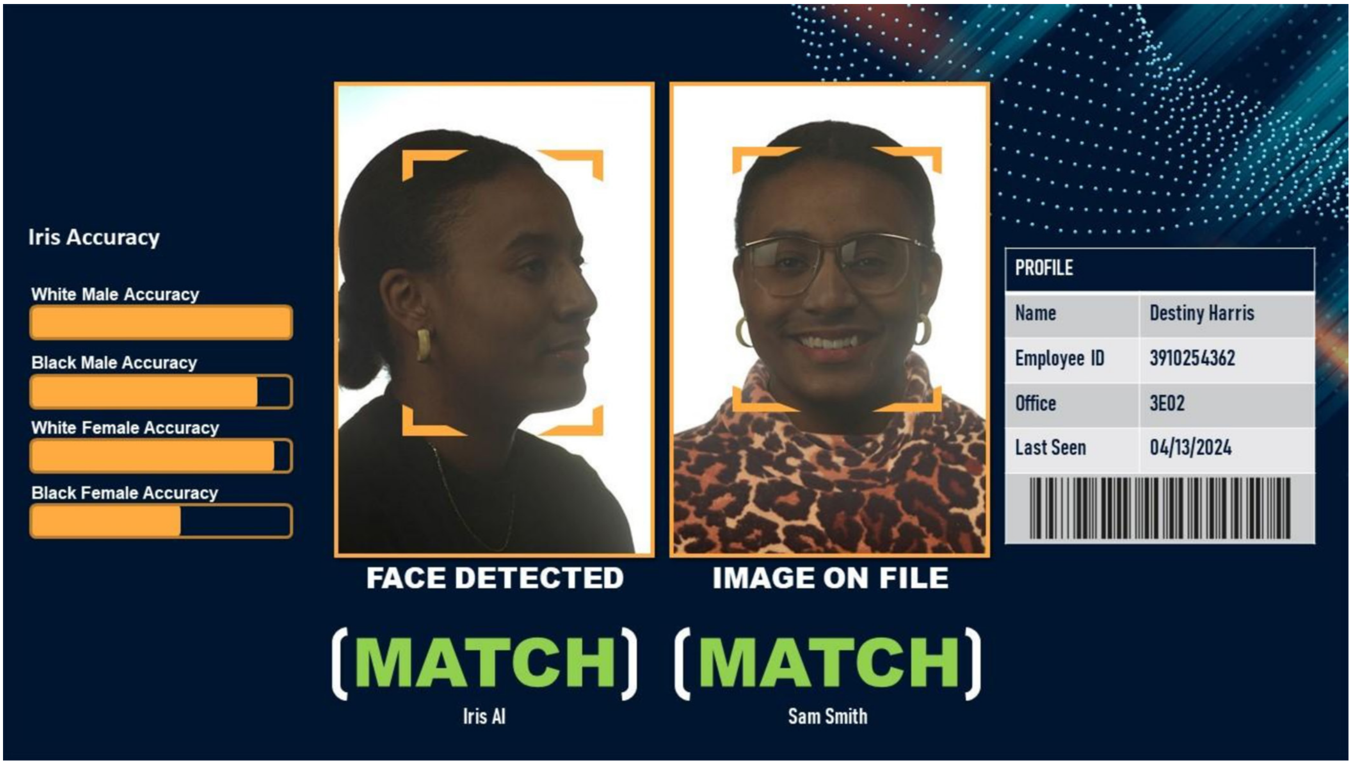


**FIGURE 3** | User interface showing congruent signal set condition in Study 3.



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

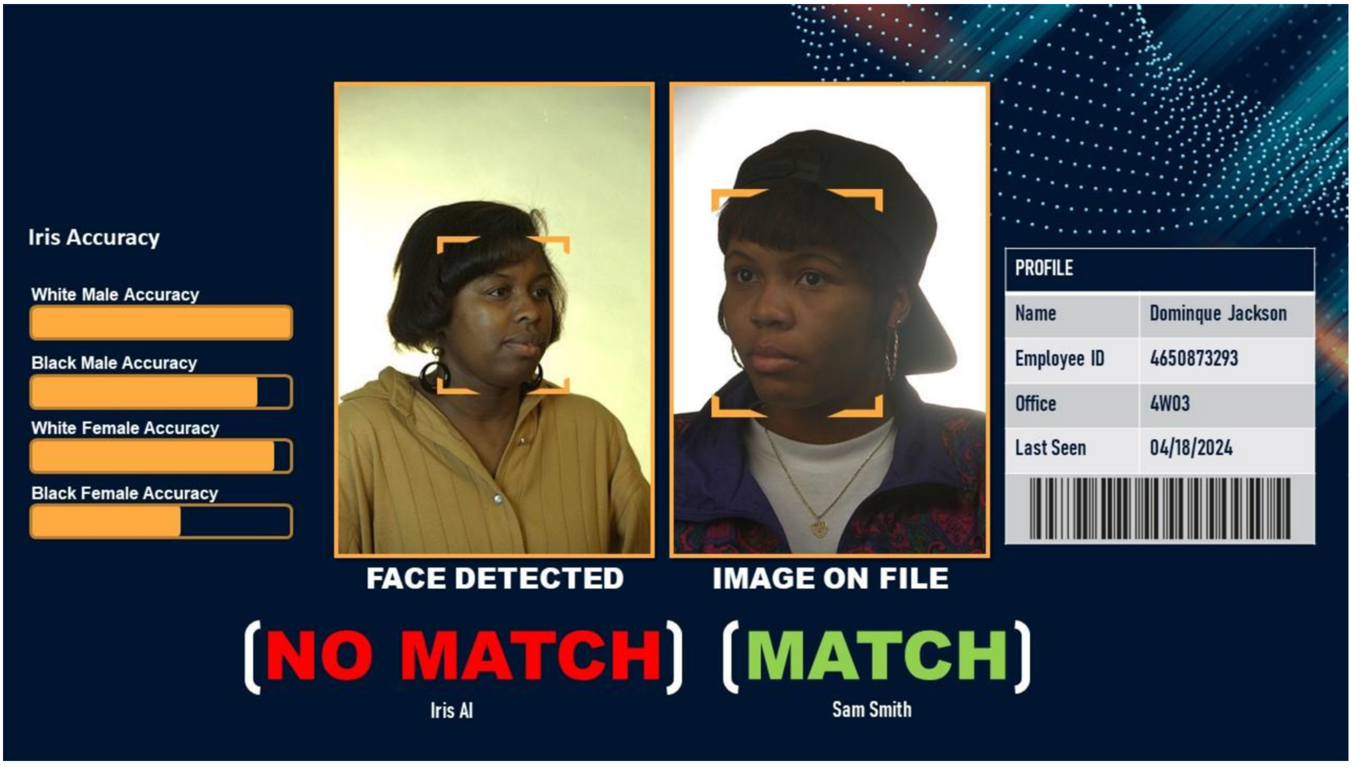


**FIGURE 4** | User interface showing incongruent signal set condition in Study 3.

**TABLE 7** | Factor loadings and retained items for algorithmic aversion scale.

| | Item | Fear of inaccuracy | Resistance to change | Loss of control | Lack of trust |
|---|---|---|---|---|---|
| 1 | I am concerned about the reliability of algorithms. | **0.648** | | | |
| 2 | Algorithms could make errors that would negatively impact my life. | **0.7582** | | | |
| 3 | I feel uncertain about the dependability of algorithm-based decisions. | **0.6559** | | | |
| 4 | The potential for inaccuracies in algorithmic results worries me. | **0.7461** | | | |
| 5 | I question the correctness of outcomes produced by algorithms. | **0.7262** | | | |
| 6 | Algorithms might not handle unexpected situations well. | **0.7465** | | | |
| 7 | I believe that algorithms can easily misinterpret complex data. | **0.737** | | | |
| 8 | The risk of error in algorithmic decisions is too high for comfort. | **0.6523** | | | |
| 9 | Algorithms may not always provide consistent results. | **0.6921** | | | |
| 10 | I am skeptical about the accuracy of algorithms when they process large amounts of data. | **0.6862** | | | |
| 11 | The thought of relying solely on algorithmic accuracy makes me uneasy. | **0.6815** | | | |
| 12 | Algorithms are not infallible and this can lead to serious mistakes. | **0.6819** | | | |
| 13 | I distrust algorithms when the stakes of the decision are high. | **0.6219** | | | |
| 14 | Algorithms could generate incorrect results without any warning signs. | **0.7461** | | | |

(Continues)



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 7 | (Continued)

| | Item | Fear of inaccuracy | Resistance to change | Loss of control | Lack of trust |
|---|---|---|---|---|---|
| 15 | I am wary of using algorithms for decisions that have long-term consequences. | **0.6519** | | | |
| 16 | I do not trust algorithms to understand my individual needs. | | | | **0.4031** |
| 17 | Algorithms lack the human insight necessary for many decisions. | | | | 0.4067 |
| 18 | I find it hard to trust a process I cannot easily understand or see. | | | | **0.6** |
| 19 | I am not confident in algorithms to handle tasks that require emotional sensitivity. | | | | 0.4881 |
| 20 | The impersonal nature of algorithms makes them hard to trust. | | | | **0.6629** |
| 21 | I feel that algorithms do not have the capability to adapt to new or evolving situations. | | | | **0.5061** |
| 22 | Algorithms are not transparent enough for me to trust their decisions. | | | | **0.5947** |
| 23 | I worry about biases embedded in algorithms affecting their judgments. | 0.4047 | | | 0.4473 |
| 24 | Trusting algorithms with personal information makes me uncomfortable. | | | | **0.4584** |
| 25 | I am skeptical about the fairness of algorithmic decisions. | | | | **0.5247** |
| 26 | Algorithms do not provide the rationale for their decisions, which diminishes my trust. | | | | **0.4482** |
| 27 | I am concerned that algorithms prioritize efficiency over ethical considerations. | | | | |
| 28 | Algorithms cannot be held accountable in the same way humans can. | | | | |
| 29 | I distrust algorithms because I cannot negotiate or reason with them. | | | | |
| 30 | The idea of algorithms making decisions without human oversight is troubling to me. | | | | |
| 31 | Relying on algorithms makes me feel like I am losing control over decisions. | | | **0.5915** | |
| 32 | I prefer to keep important decision-making processes directly under human control. | | | | |
| 33 | Algorithms taking over tasks leave me feeling helpless. | | | **0.6165** | |
| 34 | I am uneasy about delegating critical decisions to algorithms. | | | | |
| 35 | I believe that excessive use of algorithms can erode personal autonomy. | | | | |
| 36 | Algorithms decide too much, too quickly, reducing my control. | | | **0.6547** | |
| 37 | I am concerned about becoming overly dependent on algorithms. | | | **0.6418** | |
| 38 | Using algorithms feels like putting my fate in the hands of a machine. | | | **0.5138** | |
| 39 | I prefer decisions made by people because it keeps control within human hands. | | 0.438 | | |
| 40 | The use of algorithms in decision-making processes makes me feel disconnected. | | | **0.6046** | |

(Continues)



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

| | Item | Fear of inaccuracy | Resistance to change | Loss of control | Lack of trust |
|---|---|---|---|---|---|
| 41 | I think that algorithms can make choices that aren't in my best interest. | | | | |
| 42 | Algorithms reduce the human touch in decisions that affect me. | | | | |
| 43 | I worry about the consequences of errors when control is ceded to algorithms. | 0.4743 | | | |
| 44 | I am reluctant to accept decisions made without human involvement. | | | | |
| 45 | The shift toward algorithmic decision-making diminishes individual influence. | | | **0.4004** | |
| 46 | I am hesitant to adopt new technologies that rely heavily on algorithms. | | **0.6497** | | |
| 47 | I prefer traditional methods over algorithmic solutions. | | **0.8141** | | |
| 48 | The pace of change toward more algorithm use is concerning. | | **0.622** | | |
| 49 | I am skeptical about replacing human roles with algorithms. | | **0.5336** | | |
| 50 | I prefer the way things were done before algorithms were involved. | | **0.8686** | | |
| 51 | Rapid technological changes involving algorithms are unsettling. | | **0.6641** | | |
| 52 | I resist changes that involve using algorithms in my daily activities. | | **0.7925** | | |
| 53 | I believe that some things are better left unchanged, especially regarding algorithm use. | | **0.6663** | | |
| 54 | The movement toward algorithm-driven processes is too fast for my comfort. | | **0.7089** | | |
| 55 | I am cautious about the increasing reliance on algorithms in professional settings. | | **0.6269** | | |
| 56 | I prefer to stick to methods that have proven successful over time rather than adopting new algorithmic techniques. | | **0.8912** | | |
| 57 | I am wary of changes that could make algorithms central to my life or work. | | **0.7064** | | |
| 58 | I am not ready to embrace the widespread use of algorithms. | | **0.8122** | | |
| 59 | I find it difficult to accept the shift from human expertise to algorithmic control. | | **0.8074** | | |
| 60 | The trend toward automating more with algorithms does not sit well with me. | | **0.7914** | | |

*Note:* Items that cross loaded on a 5th factor were not retained in the final version of the scale. Items with bold factor loadings were in the final measure.

was adequate ($\alpha$ Algorithmic Aversion = 0.98; $\alpha$ Fear of Inaccuracy = 0.96, $\alpha$ Lack of Trust = 0.92, $\alpha$ Fear of Loss of Control = 89, and $\alpha$ Resistance to Change = 0.96).

### 4.7.3 | Manipulation Checks

Like in Studies 1 and 2, $\chi^2$ tests revealed that participants accurately recalled whether they had the option to solicit advice ($\chi^2 = 72.638$, $p < 0.000$) and the signal fit ($\chi^2 = 117.068$, $p < 0.000$). Regressions also showed that only the variable capturing each experimental condition was a significant predictor of recalling information about that condition ($p < 0.001$ in each case). To check whether our congruent signal set manipulation impacted participants, we regressed the impact of signal set congruence on participant confidence in their final decision. Mixed effects regression showed the only manipulated variable predicting final decision confidence was signal set congruence ($\boldsymbol{\gamma} = 8.449$, $p < 0.001$). The main effects of signal fit ($\boldsymbol{\gamma} = -0.734$, $p = 0.650$) and optional advice ($\boldsymbol{\gamma} = 1.309$,



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 8** | Study 3 means, correlations, and standard deviations.

| | Variable | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | AI convergence | 0.680 | 0.460 | 1.000 | | | | | | | | | | | |
| 2 | Weak signal fit | 0.480 | 0.500 | −0.021 | 1.000 | | | | | | | | | | |
| 3 | Optional AI advice | 0.530 | 0.490 | 0.019 | 0.032 | 1.000 | | | | | | | | | |
| 4 | Congruent signal set | 0.500 | 0.500 | 0.412 | 0.000 | 0.000 | 1.000 | | | | | | | | |
| 5 | Decline to solicit | 0.980 | 0.290 | −0.153 | 0.001 | 0.307 | −0.009 | 1.000 | | | | | | | |
| 6 | Initial decision confidence | 72.230 | 20.410 | −0.126 | 0.004 | −0.007 | −0.017 | 0.243 | 1.000 | | | | | | |
| 7 | Female | 47.800 | 0.500 | 0.003 | 0.022 | −0.141 | 0.000 | −0.017 | 0.025 | 1.000 | | | | | |
| 8 | Black participant | 0.490 | 0.500 | 0.008 | 0.023 | −0.116 | 0.000 | 0.009 | 0.139 | 0.009 | 1.000 | | | | |
| 9 | Explicit racism | 2.650 | 1.320 | −0.030 | 0.033 | 0.004 | 0.000 | 0.052 | 0.049 | −0.302 | 0.085 | 1.000 | | | |
| 10 | Explicit sexism | 3.200 | 1.450 | 0.010 | −0.024 | −0.096 | 0.000 | −0.005 | 0.055 | −0.284 | 0.010 | 0.550 | 1.000 | | |
| 11 | Trust in AI system | 3.650 | 0.790 | 0.084 | 0.030 | 0.061 | 0.000 | 0.062 | 0.155 | −0.058 | 0.150 | 0.158 | 0.124 | 1.000 | |
| 12 | Algorithmic aversion | 4.600 | 1.170 | −0.064 | −0.005 | 0.046 | 0.000 | −0.034 | −0.035 | 0.074 | −0.112 | 0.010 | −0.085 | −0.407 | 1.000 |

*Note: N* Level 2 = 299; *N* Level 1 = 4784. All correlations above 0.025 in absolute value are significant at $p < 0.05$.

$p = 0.419$) were not significant nor was the three-way interaction term between experimental manipulations ($\gamma = 0.718$, $p = 0.682$).

## 4.8 | Study 3 Results

Table 8 displays the means, standard deviations, and descriptive statistics of our Study 3 variables. For the third time, results show that for signal fit there was no significant effect on the likelihood of AI convergence (*Odds ratio* = 0.875, $p = 0.267$). Unlike in Studies 1 and 2, for optional AI advice, results including only main effects (Model 1) do not show a significant, negative effect on the likelihood of convergence when AI advice was optional versus when it was not optional (*Odds ratio* = 1.112, $p = 0.378$). Instead, the key driver of AI convergence is the congruence of the signal set within the team (Hypothesis 3). When the signals sent by Iris AI and Sam Smith were congruent, participants were much more likely to take the AI advice, and this effect was large (*Odds ratio* = 9.098, $p < 0.001$).

When one accounts for the choice to see the advice by both team members, Sam Smith and Iris AI, however, results again show the same pattern as the previous two studies. The impact of choosing not to see team members' advice decreases the likelihood that participants make the same decision as Iris AI (*Odds ratio* = 0.212, $p < 0.001$), while being in the optional AI condition showed a positive effect after accounting for this choice by participants (*Odds ratio* = 1.519, $p < 0.001$). Thus, Hypothesis 1 was not supported, and Hypothesis 2 was supported for the third time. Furthermore, Hypothesis 3 was supported.

### 4.8.1 | Study 3 Supplementary Analysis

Although not hypothesized, to fully probe our experimental design, we tested the three-way interaction between signal fit, optional advice, and signal set congruence on the likelihood of AI convergence. This interaction term was significant (*Odds ratio* = 1.911, $p = 0.043$). As shown in Figure 5, the largest impact of our manipulated variables is caused by the signal set

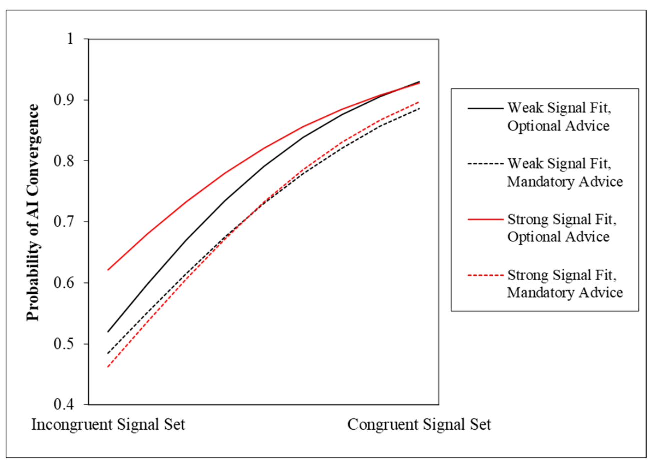


**FIGURE 5** | Interaction between optional team member advice, signal fit, and signal set congruence for Study 3.



1091379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

congruence. When the signals between Sam Smith and Iris AI are the same, participants are much more likely to take AI advice. Visual inspection suggests that this is slightly more likely when the signal is optional compared with when it is mandatory. When the signals do not align, however, there are greater differences across the optional advice and signal fit conditions. Specifically, when signal sets are not congruent, AI convergence is more likely when the signal fit is strong (when AI accuracy is not presented) and when seeing team members' advice is optional. Combined, this provides further support for Hypotheses 2 and 3 and, for the first time, suggests that signal fit may have a nuanced effect that is conditional on the presence of a human team member.

Similar to Studies 1 and 2, we controlled for participants' level of confidence and other variables that would lend support for our theorizing. Results of this analysis again show no qualitative difference in our tests of Hypotheses 1–3 when controlling for initial decision confidence, trust in AI, racial bias, and gender bias (see Tables 9 and 10). Furthermore, similar to Studies 1 and 2, participants' trust in AI significantly increased employees' AI advice taking in the team. Finally, unlike in Studies 1 and 2, explicit racism significantly decreased the likelihood that participants would take AI advice.

## 4.9 | Study 3 Discussion

The results of Study 3 replicated Hypothesis 2 for the third time testing for the additional effects of signal set congruence in human–AI teams. Although agreement between the human and AI team members was the largest driver of AI advice taking, optional AI advice again seemed to improve the likelihood that employees take AI advice, but only when participants chose to see team members' advice. Unlike in our previous studies, the three-way interaction also suggests that strong versus weak signal fit may have an effect on taking advice from AI; however, this effect is most pronounced when the signal set is incongruent and participants make a choice to see the advice of their team members.

One potential limitation of the studies presented thus far is their focus on a specific context: the facial recognition task, where AI and human partners work together to determine a definitive correct answer. This setup does not fully capture the complexities of real-world human–AI teaming, where decisions often involve ambiguity, conflicting priorities, or subjective judgments. To address this and test our hypotheses in a broader context, Study 4 examines human–AI teaming in the recruitment domain.

## 4.10 | Study 4 Method

### 4.10.1 | Participants

A total of 301 employees followed the link to participate in Study 4 and were paid 3.80 pounds in exchange for their time. We removed 42 who failed to provide consent or failed to recall basic information about the study (process checks). This resulted in 259 participants who were included in the final analyses. All participants resided in the United States, were employed

**TABLE 9** | Study 3 results.

| | Model 1 main effects only | | Model 2 main effects plus solicit decision | | Model 3 two-way interaction plus solicit decision | | Model 4 three-way interaction plus solicit decision | |
|---|---|---|---|---|---|---|---|---|
| **Variable** | **Odds ratio** | **SE** | **Odds ratio** | **SE** | **Odds ratio** | **SE** | **Odds ratio** | **SE** |
| Weak signal fit | 0.87 | 0.10 | 0.87 | 0.10 | 1.02 | 0.17 | 1.09 | 0.20 |
| Optional AI advice | 1.11 | 0.13 | 1.52** | 0.18 | 1.76** | 0.29 | 1.91** | 0.34 |
| Congruent signal set | 9.09** | 0.74 | 9.51** | 0.79 | 9.51** | 0.79 | 10.16** | 1.64 |
| Weak signal fit × Optional advice | | | | | 0.74 | 0.17 | 0.60* | 0.15 |
| Weak signal fit × Congruent signal set | | | | | | | 0.81 | 0.19 |
| Optional advice × Congruent signal set | | | | | | | 0.77 | 0.17 |
| Weak signal fit × Optional advice × Congruent signal set | | | | | | | 1.91* | 0.61 |
| Decline to solicit | | | 0.21** | 0.03 | 0.21** | 0.03 | 0.21** | 0.03 |
| Intercept | 0.96 | 0.10 | 0.94 | 0.10 | 0.87 | 0.10 | 0.86 | 0.11 |
| var (participant ID) | 0.68 | 0.10 | 0.59 | 0.09 | 0.59 | 0.09 | 0.59 | 0.09 |

**$p < 0.01$.
*$p < 0.05$.
†$p < 0.1$.



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 10** | Study 3 supplementary analysis results.

| Variable | Model 1 main effects only | | Model 2 main effects plus solicit decision | | Model 3 two-way interaction plus solicit decision | | Model 4 three-way interaction plus solicit decision | |
|---|---|---|---|---|---|---|---|---|
| | Odds ratio | SE | Odds ratio | SE | Odds ratio | SE | Odds ratio | SE |
| Weak signal fit | 0.87 | 0.10 | 0.87 | 0.10 | 1.05 | 0.17 | 1.13 | 0.20 |
| Optional AI advice | 1.13 | 0.14 | 1.49** | 0.18 | 1.78** | 0.29 | 1.96** | 0.34 |
| Congruent signal set | 9.44** | 0.78 | 9.72** | 0.81 | 9.72** | 0.81 | 10.50** | 1.70 |
| Weak signal fit × Optional AI advice | | | | | 0.69 | 0.15 | 0.56* | 0.14 |
| Weak signal fit × Congruent signal set | | | | | | | 0.81 | 0.19 |
| Optional advice × Congruent signal set | | | | | | | 0.74 | 0.17 |
| Weak signal fit × Optional advice × Congruent signal set | | | | | | | 1.99** | 0.64 |
| Decline to solicit | | | 0.27** | 0.04 | 0.27** | 0.04 | 0.27** | 0.04 |
| Initial confidence in decision | 0.98** | 0.002 | 0.98 | 0.002 | 0.98 | 0.002 | 0.98** | 0.002 |
| Female participant | 1.07 | 0.13 | 1.11 | 0.13 | 1.10 | 0.13 | 1.10 | 0.13 |
| Black participant | 1.13 | 0.14 | 1.14 | 0.13 | 1.14 | 0.13 | 1.14 | 0.13 |
| Explicit racial bias | 0.88* | 0.05 | 0.89* | 0.05 | 0.88* | 0.05 | 0.89** | 0.05 |
| Explicit gender bias | 1.08 | 0.05 | 1.08 | 0.05 | 1.08† | 0.05 | 1.08† | 0.05 |
| Trust in AI system | 1.44** | 0.12 | 1.43** | 0.11 | 1.44** | 0.11 | 1.44** | 0.11 |
| Algorithmic aversion | 0.92 | 0.05 | 0.91† | 0.05 | 0.90† | 0.05 | 0.91† | 005 |
| Intercept | 1.88 | 0.94 | 1.33 | 0.63 | 1.19 | 0.56 | 1.15 | 0.55 |
| var (participant ID) | 0.59 | 0.09 | 0.51 | 0.08 | 0.50 | 0.08 | 0.51 | 0.08 |

**$p < 0.01$.
*$p < 0.05$.
†$p < 0.1$.

full-time or part-time, held a leadership position, and had experience in hiring employees in the United States. Participants were on average 42.53 years old (SD 12.49). Approximately 49.8% of our participants identified as women, 48.6% of participants identified as men, and 1.4% identified as transgender (half transgender men, half transgender women). Approximately 33.2% of participants identified as Black/Caribbean/African/Afro Latin, 54.4% identified as White/Caucasian/European, 5% identified as Asian/Indian/Southeast Asian, 5% identified as Hispanic/Latino/Latinx, non-White, 0.4% identified as American Indian/Alaska Native/Indigenous, 1.5% identified as Mixed/Multiple Identity, and 0.4% identified as Other.

#### 4.10.2 | Procedure

In Study 4, we adapted the experimental design from Study 3 to the recruitment context. Each participant was randomly assigned into one of four experimental conditions following a 2 (signal fit: strong versus weak) × 2 (option to solicit advice from team members: yes or no) between-subject design. Participants were provided a job description for a junior software position and were given the task of evaluating eight CVs for a software engineering position. Within each condition, participants completed a series of eight CV evaluation tasks. Participants were instructed to play the role of the leader of a team consisting of a human coworker Sam Smith and a resume screening algorithm, Iris AI. Participants' task was to choose whether to invite each candidate to an interview. As the leader of the human–AI team, they held the final decision-making authority.

We developed eight fictional CVs that differed by gender and race of the applicant while keeping professional qualifications (education, previous work experience, and hard and soft skills) directly comparable. Half of the fictional CVs featured female



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and racial minority applicants. As the study focused on the United States, we chose Black/African American–sounding names to signal the ethnic background of the candidates. Gender and ethnicity could be inferred from the names (e.g., Anna Wagner and Xavier Jeanbaptiste). Figure 6 presents an example of a fictional CV. All job applicants held a bachelor's degree

**Anna Schmidt**

*Application Details for Software Engineering position*

*All sections are required except where noted. For candidates who are interviewed, all information entered below will be verified.*

**Work Experience**

*Title*: Software Engineer
*Company Name*: Google
*Location*: Mountain View, CA
*Dates*: 10/2020 - Current position

***Description, Duties, Responsibilities***:
- Focusing on working in and enhancing the whole lifecycle of services - to design, develop, test, deploy, maintain, and improve software.
- Researching, conceiving, and developing web applications to extend and improve Google's product offering.
- Collaborating on scalability issues involving massive amounts of data and information.

*Title: Software Engineering Intern*
*Company Name*: IBM
*Location*: Armonk, NY
*Dates*: 05/2019 - 11/2019

***Position Description, Duties, Responsibilities:***
- Supported the engineering team at IBM by providing software development expertise for various projects, contributing to the enhancement of IBM's products and services.
- Implemented code to handle large volumes of data related to IBM's operations, ensuring scalability, efficiency, and reliability.
- Participated in code reviews, testing, and debugging activities to ensure the quality and stability of software solutions.
- Contributed to the continuous improvement of development processes and methodologies, identifying areas for optimization and efficiency gains.

**Education**

*University/college*: Cornell University
*Location*: Ithaca, NY
*Dates*: 2016 - 2020
*Graduated:* Yes
*Degree:* BS (Bachelor of Science)
*Subject/Major:* Computer Science
*Relevant Coursework: Operating systems, Parallel computer architecture, Data structures and algorithms, Computer networks, Databases*

**Additional Skills**

Sawzall, C/C++, Java, Javascript, Karma, Python, R, Perl

**FIGURE 6** | Example CV from Study 4.



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in software engineering from one of the Top 10 universities in the United States awarded between 2017 and 2020. Additionally, they completed an internship of 3–6 months and accumulated 3–5 years of work experience with one of the Top 10 employers in the United States within their respective fields. The selection of applicants' names, universities, and employers was meticulously drawn from reliable sources, including top baby name lists in the 1990s, top employers' rankings, and prestigious university rankings in the United States.

The experimental setup followed a two-step process parallel to that of Studies 1–3. In the first step, participants were asked to review the CV of a job candidate and to decide whether to invite them to an interview. In the second step, depending on their condition, participants were sent a signal with strong versus weak fit from their teammates (Sam Smith and Iris AI) and were either required to see their teammates' decisions (mandatory advice) or were allowed to choose whether to see their teammates' advice (optional advice). Participants then had to make a final decision whether to invite the candidate for an interview. In total, participants evaluated a series of eight pretested CVs that were evenly balanced by race (Black vs. White names) and gender (women's vs. men's names).

Each employee was given signals in a format similar to Studies 1–3 (see Figures 7 and 8). The signal fit was manipulated using similar bar charts as in the previous studies. In the weak signal fit condition, participants saw Iris AI's historical selection ratio of applicants from each of the four demographic categories (Black vs. White applicants) and (women vs. men applicants). For half of the CVs, Sam Smith provided the same advice as Iris AI, whereas for the other half, they provided the opposite advice. These recommendations were evenly balanced between "yes, invite for an interview" and "no, do not invite for an interview" decisions. To mirror the high-stakes scenario of previous studies, participants were informed that the stakes were high for this open position and that they could interview only four out of the eight job candidates.

### 4.10.3 | Measures

The same focal measures that were used in Study 3 were used in Study 4. The internal consistency of our control measures was adequate ($\alpha$ Trust in CV Screening Algorithms = 0.94; $\alpha$ Explicit Gender Bias = 0.8; $\alpha$ Explicit Racial Bias = 0.74; Algorithm Aversion = $\alpha$ 0.98). Given we had created a new measure of algorithmic aversion for Study 3, we assessed the measurement model of our new measure using confirmatory factor analysis in Stata 16 using the SEM command. Comparative fit index (CFI) values below 0.90 indicate poor model fit, 0.90–0.95 indicate marginal fit, and above 0.95 indicate the model demonstrates good fit. Further, RMSEA and SRMR values above 0.09 indicate poor fit, 0.08–0.06 indicate marginal fit, and below 0.06 indicate good fit (Hu and Bentler 1998). Our model specifying one higher order dimension that consisted of four subdimensions demonstrated adequate fit (CFI = 0.911, RMSEA = 0.070, SRMR = 0.047).

### 4.10.4 | Manipulation Checks

At the end of Study 4, we asked participants to choose whether they "always had information available from their teammates," "had no information from their teammates," or "had information available from their teammates, but could choose not to see it." $\chi^2$ tests revealed that participants accurately recalled whether they had the option to solicit information from teammates ($\chi^2 = 435.57$, $p < 0.001$). Likewise, the check for signal fit worked as expected ($\chi^2 = 7.64$, $p = 0.006$). To check whether our

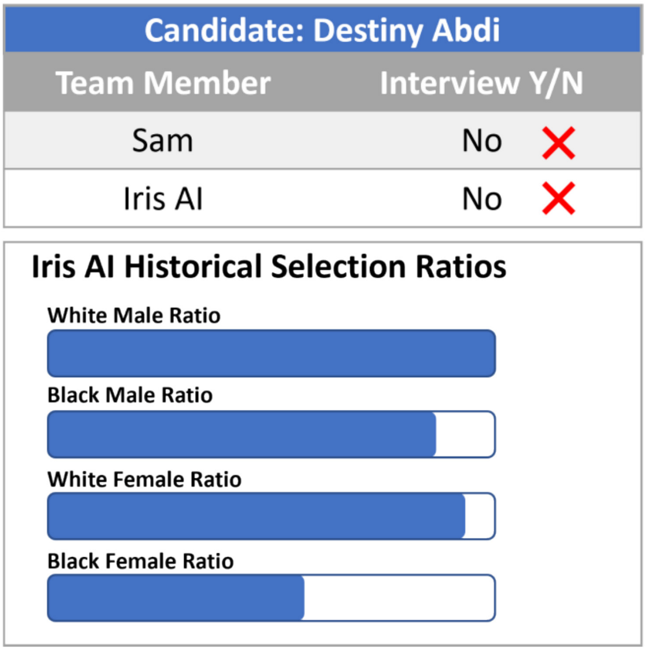


**FIGURE 7** | Example user interface showing congruent signal set condition in Study 4. *Note:* In the strong signal fit condition, participants were only shown the top table with team member decisions.

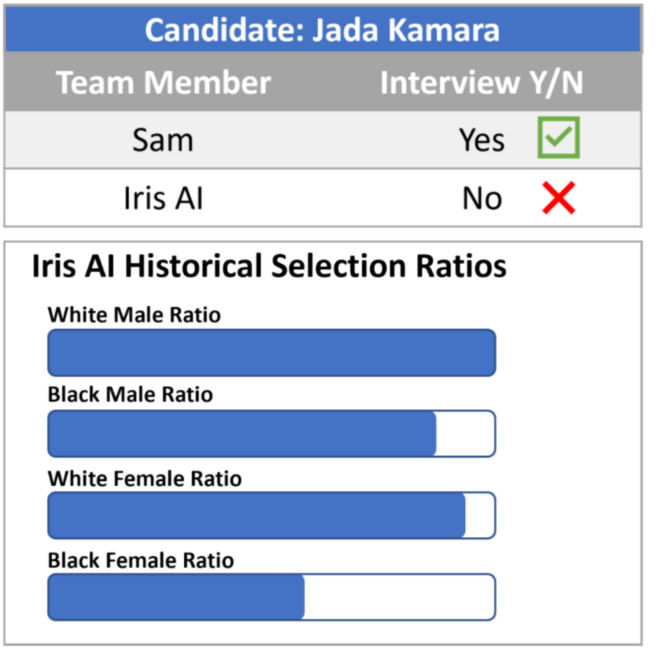


**FIGURE 8** | Example user interface showing incongruent signal set condition in Study 4. *Note:* In the strong signal fit condition, participants were only shown the top table with team member decisions.



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

congruent signal manipulations impacted participants, we regressed the impact of signal congruence on participant confidence in their final decision. Mixed effects regression including only the main effects of our experimental conditions shows that the only variable predicting final decision confidence was the signal congruence variable ($\gamma = 3.109$, $p < 0.001$). Looking at interactions, the main effects of signal fit ($\gamma = 0.095$, $p = 0.970$) and optional advice ($\gamma = -0.722$, $p = 0.773$) were not significant. The three-way interaction between experimental manipulations was trending, though not significant ($\gamma = -4.759$, $p = 0.065$).

## 4.11 | Study 4 Results

Table 11 displays the means, standard deviations, and descriptive statistics of our Study 4 variables. Results show that when accounting for only the main effects of our experimental conditions, signal fit had no significant effect on the likelihood of AI convergence regardless of whether additional information was presented as bar charts showing historical selection ratios of Iris AI (*Odds ratio* = 0.879, $p = 0.183$). Furthermore, like in previous studies, we found that being in the optional AI advice condition reduced the likelihood of convergence between participants and Iris AI (*Odds ratio* = 0.672, $p < 0.001$), replicating the findings from Studies 1–3 related to Hypothesis 2. As in Study 3, we find the key driver of AI convergence is the congruence of the signal set within the team (Hypothesis 3). When the signals sent by Iris AI and Sam Smith were congruent, participants' decisions were much more likely to converge with Iris AI (*Odds ratio* = 3.717, $p < 0.001$).

Accounting for the choice to see the advice by Sam Smith and Iris AI in the optional AI conditions, we find a different pattern of results than was found in Studies 1–3. First, the effect of choosing to ignore teammates' advice decreased the likelihood of taking Iris AI's advice (the direction of the effect was the same as in previous studies), but this effect was only trending (*Odds ratio* = 0.724, $p = 0.085$). Furthermore, the impact of being in the optional AI condition did not become positive when accounting for participants' decisions to see team members' advice. In other words, the effect remained negative and significant (*Odds ratio* = 0.701, $p < 0.001$). Thus, Hypothesis 1 was not supported, and Hypothesis 2 was only weakly supported. Hypothesis 3 was supported.

### 4.11.1 | Study 4 Supplementary Analysis

The interaction term between signal fit, optional advice, and signal congruence was not significant (*Odds ratio* = 0.891, $p = 0.774$). Similar to Studies 1–3, we controlled for the same variables as in previous studies. Results of this analysis again show no qualitative difference in our tests of Hypotheses 1–3 when adding the controls (see Tables 12 and 13). Interestingly, identifying as a Black participant significantly reduced the likelihood of AI convergence across all models.

## 5 | General Discussion

As organizations increasingly rely on human–AI teaming to enhance decision-making processes, understanding how and

**TABLE 11** | Study 4 means, correlations, and standard deviations.

| | Variable | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Took AI advice | 0.64 | 0.48 | 1.000 | | | | | | | | | | | |
| 2 | Weak signal fit | 0.53 | 0.50 | −0.024 | 1.000 | | | | | | | | | | |
| 3 | Optional AI | 0.53 | 0.50 | −0.085 | −0.039 | 1.000 | | | | | | | | | |
| 4 | Congruent signal set | 0.50 | 0.50 | 0.299 | 0.000 | 0.000 | 1.000 | | | | | | | | |
| 5 | Decline to solicit | 0.07 | 0.26 | −0.042 | 0.006 | 0.263 | 0.050 | 1.000 | | | | | | | |
| 6 | Initial decision confidence | 74.08 | 18.24 | −0.023 | 0.028 | 0.013 | −0.014 | 0.032 | 1.000 | | | | | | |
| 7 | Female | 0.50 | 0.50 | 0.002 | −0.042 | −0.042 | 0.000 | −0.101 | −0.076 | 1.000 | | | | | |
| 8 | Black participant | 0.33 | 0.47 | −0.065 | 0.069 | 0.036 | 0.000 | 0.086 | 0.104 | −0.047 | 1.000 | | | | |
| 9 | Explicit racism | 2.81 | 1.27 | −0.001 | −0.046 | 0.071 | 0.000 | 0.134 | −0.126 | −0.184 | 0.182 | 1.000 | | | |
| 10 | Explicit sexism | 3.18 | 1.38 | −0.011 | −0.061 | 0.067 | 0.000 | 0.040 | −0.052 | −0.286 | 0.104 | 0.427 | 1.000 | | |
| 11 | Trust in AI system | 3.53 | 0.91 | 0.030 | −0.086 | −0.047 | 0.000 | 0.007 | 0.066 | −0.095 | 0.342 | 0.251 | 0.157 | 1.000 | |
| 12 | Algorithmic aversion | 4.34 | 1.39 | −0.030 | 0.071 | 0.046 | 0.000 | −0.049 | −0.051 | 0.093 | −0.127 | −0.180 | −0.085 | −0.610 | 1.000 |

*Note:* $N$ Level 2 = 259; $N$ Level 1 = 2072. All correlations above 0.025 in absolute value are significant at $p < 0.05$.



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 12** | Study 4 results.

| Variable | Model 1 main effects only | | Model 2 main effects plus solicit decision | | Model 3 two-way interaction plus solicit decision | | Model 4 three-way interaction plus solicit decision | |
|---|---|---|---|---|---|---|---|---|
| | Odds ratio | SE | Odds ratio | SE | Odds ratio | SE | Odds ratio | SE |
| Weak signal fit | 0.88 | 0.09 | 0.88 | 0.08 | 0.92 | 0.13 | 0.97 | 0.18 |
| Optional advice | 0.67** | 0.07 | 0.70** | 0.07 | 0.73* | 0.11 | 0.73 | 0.13 |
| Congruent signal set | 3.71** | 0.37 | 3.76** | 0.37 | 3.76** | 0.37 | 4.44** | 1.02 |
| Weak signal fit × Optional advice | | | | | 0.93 | 0.18 | 0.99 | 0.24 |
| Weak signal fit × Congruent signal set | | | | | | | 0.83 | 0.25 |
| Optional advice × Congruent signal set | | | | | | | 0.95 | 0.28 |
| Weak signal fit × Optional advice × Congruent signal set | | | | | | | 0.89 | 0.36 |
| Decline to solicit | | | 0.72† | 0.14 | 0.73† | 0.14 | 0.73† | 0.14 |
| Intercept | 1.31 | 0.12 | 1.31 | 0.13 | 1.27* | 0.14 | 1.20 | 0.16 |
| var (participant ID) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

*Note:* $N$ Level 2 = 259; $N$ Level 1 = 2072.
**$p < 0.01$.
*$p < 0.05$.
†$p < 0.01$.

why employees align their decisions with AI is essential. This alignment has the potential to enhance decision accuracy and consistency, benefiting organizations. However, it also raises critical questions about ethics, trust, and under or over reliance on AI recommendations (Andrieux et al. 2024; Glikson and Woolley 2020). Drawing on signaling theory (Spence 1973) and self-determination theory (Deci and Ryan 1985), this research helps organizations understand what factors are and are not relevant to AI convergence in human–AI teams. Moreover, by understanding AI convergence, we can begin to uncover the complexities of human–AI teaming, including scenarios where AI divergence may be the optimal outcome. By advancing this understanding, we contribute significantly to the existing literature and pave the way for more effective AI applications in human–AI teams.

We focused our attention on three qualities of signals in human–AI teams—signal fit, whether signals were optional, and whether human team members sent signals that were congruent or incongruent to that of their AI teammate. Specifically, we predicted that signal fit would impact the likelihood of AI convergence as signals with a weak fit are less effective in reducing uncertainty. Further, we hypothesized that operating in a team where advice from AI was optional would increase AI convergence, as long as employees chose to see AI advice. We argue that employees' need for autonomy is satisfied when they have a choice to see AI advice, and their need for competence is fulfilled when they seek more comprehensive information from their AI counterparts, thereby increasing the likelihood of AI convergence. Finally, we predicted the impact of incongruent signal sets (when a human team member disagrees with an AI team member) would decrease the likelihood that employees aligned with an AI decision.

In our first three experimental studies in the facial recognition context, Hypothesis 2, which posited that the availability of optional AI advice and its active selection increases AI convergence, was supported. However, in Study 4, where participants were evaluating CVs, a task that has direct implications on the lives and careers of job candidates, we found that optional AI advice decreased the likelihood of AI convergence, and this was independent of the negative effect of choosing not to see team members' advice. This pattern of results is in line with previous research suggesting that employees are more open to working with AI when they maintain final authority over decisions and moral implications of their decisions are low (Chugunova and Sele 2022; Haesevoets et al. 2021).

In contrast, Hypothesis 1 was not significant in all four studies suggesting that the signal fit does not play a crucial role in AI convergence in human–AI teams. This finding resembles previous research on human–AI systems, covering domains like medical treatments (Jacobs et al. 2021), loan defaults (Green and Chen 2019), and income prediction (Zhang, Liao, and Bellamy 2020). Although there is an anticipation that providing explanations for AI decisions (i.e., increasing or decreasing signal fit) would impact teams by aiding human understanding and error identification, recent findings suggest that such explanations often fail to help humans utilize the information effectively for making better decisions (Buçinca, Malaya, and



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 13** | Study 4 supplementary analysis results.

| Variable | Model 1 main effects only | | Model 2 main effects plus solicit decision | | Model 3 two-way interaction plus solicit decision | | Model 4 three-way interaction plus solicit decision | |
|---|---|---|---|---|---|---|---|---|
| | Odds ratio | SE | Odds ratio | SE | Odds ratio | SE | Odds ratio | SE |
| Weak signal fit | 0.92 | 0.09 | 0.92 | 0.09 | 0.94 | 0.14 | 1.00 | 0.18 |
| Optional AI advice | 0.69** | 0.07 | 0.72** | 0.92 | 0.73* | 0.11 | 0.73 | 0.13 |
| Congruent signal set | 3.75** | 0.37 | 3.79** | 0.37 | 3.79** | 0.37 | 4.47** | 1.03 |
| Weak signal fit × Optional AI advice | | | | | 0.96 | 0.19 | 1.03 | 0.26 |
| Weak signal fit × Congruent signal set | | | | | | | 0.83 | 0.26 |
| Optional advice × Congruent signal set | | | | | | | 0.95 | 0.29 |
| Weak signal fit × Optional advice × Congruent signal set | | | | | | | 0.88 | 0.36 |
| Decline to solicit | | | 0.74 | 0.14 | 0.74 | 0.14 | 0.75 | 0.14 |
| Initial confidence in decision | 0.99 | 0.002 | 0.99 | 0.002 | 0.99 | 0.002 | 0.99 | 0.002 |
| Female participant | 0.99 | 0.10 | 0.97 | 0.10 | 0.98 | 0.10 | 0.97 | 0.10 |
| Black participant | 0.69** | 0.08 | 0.69** | 0.08 | 0.69** | 0.08 | 0.69** | 0.08 |
| Explicit racial bias | 1.01 | 0.04 | 1.02 | 0.05 | 1.02 | 0.05 | 1.02 | 0.05 |
| Explicit gender bias | 0.98 | 0.04 | 0.98 | 0.04 | 0.98 | 0.04 | 0.98 | 0.04 |
| Trust in AI system | 1.11† | 0.08 | 1.09 | 0.08 | 1.09 | 0.08 | 1.10 | 0.08 |
| Algorithmic aversion | 0.99 | 0.04 | 0.98 | 0.04 | 0.98 | 0.04 | 0.98 | 0.04 |
| Intercept | 1.08 | 0.54 | 1.10 | 0.56 | 1.09 | 0.56 | 1.03 | 0.53 |
| var (participant ID) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

*Note:* $N$ Level 2 = 259; $N$ Level 1 = 2072.
** $p < 0.01$.
* $p < 0.05$.
† $p < 0.1$.

Gajos 2021; Sergeeva et al. 2023; Vaccaro, Almaatouq, and Malone 2024).

Hypothesis 3 focused on the congruence of signal sets, when multiple signals are received from AI and human team members. We find that AI convergence is significantly more likely when signals give rise to uniform congruence compared with incongruent signal sets. This finding was replicated in both facial recognition and hiring contexts. After interpreting incongruent signals from both AI and fellow human teammates, employees tend to exhibit human convergence rather than AI convergence. However, our three-way interaction in Study 3 suggests that this pattern may vary depending on factors such as the type of task, signal fit, and whether AI advice is optional.

## 5.1 | Theoretical Implications

Building on prior research, our study investigates human–AI teaming within various organizational contexts (Lundberg et al. 2018; Patel et al. 2019), specifically extending our knowledge of the factors influencing AI convergence in human–AI teams. In particular, our research applies signaling theory coupled with self-determination theory to explore how qualities of specific signals, such as signal fit, optional AI advice, and signal set congruence impact employees' AI convergence. Although signaling theory has already been applied to a variety of different contexts (e.g., entrepreneurship, leadership, and diversity), to our knowledge, we are the first to utilize it to study teams. Understanding how employees interpret and respond to AI signals from multiple sources helps to improve the design and



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

implementation of these teams in organizations. The expansion of the signaling theory offers new avenues for research and encourages interdisciplinary approaches that help shed light on the impact of signals on the AI convergence in different contexts. Further, incorporating facial recognition systems as a novel context into the field of organizational behavior, this study explores unique high-stakes scenarios in the complex and uncertain work environment.

Two main findings stand out in our studies and contribute to the theory. First, optional AI advice is an important element of human–AI teaming. When employees choose not to see AI team members' advice, they are much less likely to make the same decision as their AI teammate. This rather intuitive finding becomes more compelling with our results indicating that after accounting for this behavior by employees, AI convergence was most likely when employees had optional AI advice that they chose to see. This corroborates results from previous studies suggesting that humans appreciate AI advice and are willing to rely on it as long as they retain the ultimate control over the decision-making process (Bundorf, Polyakova, and Tai-Seale 2019; Dietvorst, Simmons, and Massey 2018; Haesevoets et al. 2021; Logg, Minson, and Moore 2019). Furthermore, our findings lend credibility to the idea that self-determination theory is important in understanding the impact of signals in human–AI teams. As human–AI teams become commonplace, AI convergence will be impacted by work design and team design to the extent they foster or inhibit autonomy, competence, and relatedness (Gagné et al. 2022). Our research suggests that having the option to solicit AI advice rather than making AI mandatory is one path to influencing AI convergence in teams.

Second, employees were less likely to exhibit AI convergence when their human counterpart provided an incongruent signal compared with their AI teammate. Another way of interpreting this finding is that when humans and AI on a team are misaligned in their signals, human convergence is much more likely than AI convergence. The results of exploratory analyses in Study 3 showed a significant three-way interaction between signal congruence, signal fit, and optional advice, such that the negative effect of incongruent signals on AI convergence could be somewhat mitigated when the signal fit was strong, and seeing the AI advice was optional. However, the three-way interaction was not replicated in Study 4, possibly because of the more ethically charged decision context.

In combination, these findings highlight a significant limitation in human–AI teams and suggest potential remedies that require further investigation. When disagreements, assessments, and advice are incongruent among team members, the likelihood of AI convergence diminishes, though this could potentially be mitigated through the deliberate design of AI signals and processes that prioritize human choice in decision-making.

## 5.2 | Practical Implications

Our research findings hold valuable managerial implications. First, organizations seeking to form and utilize human–AI teams for solving complex tasks should focus on encouraging a collaborative approach, where all team members' advice is taken into consideration. It is essential that AI advice is viewed as a complementary tool that employees choose to see rather than a mandatory replacement for human expertise and judgment. Second, organizations and their leaders need to ensure that employees on human–AI teams take AI advice seriously. In designing human–AI team processes, it is critical to ensure that employees are more inclined to choose to see AI advice rather than ignore it. Companies should emphasize this through employee training. Third, giving human–AI teams tasks that are likely to result in disagreements between human team members and AI team members will reduce the likelihood that AI advice is taken. In these situations, creative solutions and redundancies may be necessary. Fourth, designing a simple user interface that facilitates easy access to advice from both human and AI team members and encourages interaction among team members (Noti and Chen 2023) is a critical component for effective human–AI teaming in decision-making. Finally, given the impact of incongruent signal sets, businesses may want to introduce optional AI advice gradually into their organizational processes by starting with noncritical tasks or projects where the positive impact of AI advice can be easily assessed (Einola and Khoreva 2022).

## 5.3 | Future Research and Limitations

Our results provide empirical evidence that optional AI advice and congruent signal sets enhance AI convergence when humans elect to utilize the AI, but that signal fit is less important than we had initially believed. However, in Study 4, we did not find that optional AI advice had a positive effect on AI congruence, as participants were more likely to take AI advice when they were forced to see it rather than when it was optional. Future research should explore the boundary conditions of our predictions in both experimental and field-based research. In mastering the task of evaluating CVs, participants needed to invest more time and energy into their initial decisions before they had an opportunity to see their team members' advice. This may have influenced how participants approached optional advice. Additionally, Study 4 required participants to select interview candidates from a pool of equally qualified individuals, where no candidate was clearly unfit for the job, contrasting with Studies 1–3 that had clear, objective answers. Furthermore, in our first three studies, trust in AI was a significant predictor of AI convergence, yet this was not the case in Study 4. Although we can only speculate on what might have caused these results, we believe that future research should investigate what factors might alter the impact of optional AI advice on AI convergence. Specifically, if there is a threshold of time and effort beyond which employees become less likely to take AI advice, this would be critical for organizations to know. This is all the more important given that AI is not infallible, and in some situations, AI convergence may not be the optimal outcome. Future research should explore the conditions under which AI divergence is the optimal outcome, yet may be hindered by overreliance on AI or poorly designed decision-making processes. Similarly, it could be the type of task and its moral implications, which causes mandatory AI advice to lead to AI convergence rather than optional advice. Finally, there could be different ways of manipulating signal fit that alter the impact of optional AI advice and congruent signal sets on AI convergence.



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

It is important for researchers to replicate and extend our findings. Scholars should especially consider what additional factors contribute to employees' decision to search for and utilize optional AI advice. Understanding what drives team members to choose to see AI advice when they have the option to ignore it could help scholars understand when and how optional AI advice has the greatest impact. It is possible that different types of information and various types of explanations vary in their effectiveness in providing signal credibility and fit across contexts and among individuals. Exploring this aspect further could be a fruitful area for future research. Another promising area would be to investigate further components of signals and signal sets (e.g., response time, receiver feedback, interactions between team members, and repeated signals) that may influence AI convergence in human–AI teams. Finally, although AI does not have free will by today's technology standards, the extent to which employees perceive the AI to have warmth, competence, or agency may alter how signals sent by AI are interpreted. Future research should explore this possibility especially given research suggesting cross-cultural differences may influence how individuals attend to and interpret various signals and signal sets from AI and robots (e.g., Mantello et al. 2023).

Our conclusions must be considered in light of a few limitations. Across four studies, one of our three hypothesized relationships was not significant, and in one experiment, results, in part, were the opposite of our predictions. Although we were unable to replicate the three-way interaction from Study 3, our results do suggest that future research should continue to explore the congruence of signal sets, optional AI advice, and signal fit. A further limitation of our studies is that the hypothetical scenarios we presented to participants involved at most a team of three members. In reality, human–AI teams can be larger, and this adds complexity to the potential combinations of congruent and incongruent signal sets. We encourage scholars to investigate in the future how balanced and imbalanced signal sets, as described by Drover, Wood, and Corbett (2018), influence AI convergence.

## 6 | Conclusion

The integration of human intelligence and AI in human–AI teams is becoming indispensable in today's organizations. Despite its extensive use, there is still a limited understanding of the factors guiding teams toward AI convergence over human convergence in the workplace. This research addresses this knowledge gap by investigating the main factors influencing AI convergence across different contexts. By integrating signaling theory with self-determination theory, our research illustrates how different signal characteristics and the congruence between signal sources affect AI convergence. Our results emphasize the role of optional AI advice and congruent signals from both human and AI team members in facilitating the acceptance of AI advice.


### Acknowledgments

We are grateful for the financial support from the Joachim Herz Foundation, Germany and Max Planck Society. Open Access funding enabled and organized by Projekt DEAL.

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

The data that support the findings of this study are openly available in OSF at https://osf.io/zjwta/?view_only=74a15f71b21c4b0abb69f3e5de543537.


### Endnotes

[1] A study following the same design as Study 1 was included in the original version of this paper but was removed based on thoughtful reviewers' suggestions because of an unbalanced sample on Prolific. Although the main effects of signal fit and optional AI advice were consistent with the reported findings, they were qualified by an unexpected interaction. To be conservative, we report only a single replication of this finding, with the data and results from the removed study included in our online appendix.

### References


Adam, M., E. Croitor, D. Werner, A. Benlian, and M. Wiener. 2022. "Input Control and Its Signalling Effects for Complementors' Intention to Join Digital Platforms." *Information Systems Journal* 33, no. 3: 437–466. https://doi.org/10.1111/isj.12408.

Agrawal, A., J. S. Gans, and A. Goldfarb. 2019. "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction." *Journal of Economic Perspectives* 33, no. 2: 31–50. https://doi.org/10.1257/jep.33.2.31.

Ali, S., T. Abuhmed, S. El-Sappagh, et al. 2023. "Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence." *Information Fusion* 99: 101805. https://doi.org/10.1016/j.inffus.2023.101805.

Allison, T. H., A. F. McKenny, and J. C. Short. 2013. "The Effect of Entrepreneurial Rhetoric on Microlending Investment: An Examination of the Warm-Glow Effect." *Journal of Business Venturing* 28, no. 6: 690–707. https://doi.org/10.1016/j.jbusvent.2013.01.003.

Altintas, O., A. A. Seidmann, and B. Gu. 2023. "The Effect of Interpretable Artificial Intelligence on Repeated Managerial Decision-Making Under Uncertainty." SSRN. https://doi.org/10.2139/ssrn.4331145.

Andrieux, P., R. D. Johnson, J. Sarabadani, and C. Van Slyke. 2024. "Ethical Considerations of Generative AI-Enabled Human Resource Management." *Organizational Dynamics* 53, no. 1: 101032. https://doi.org/10.1016/j.orgdyn.2024.101032.

Appels, M. 2022. "CEO Sociopolitical Activism as a Signal of Authentic Leadership to Prospective Employees." *Journal of Management* 49, no. 8: 2727–2765. https://doi.org/10.1177/01492063221110207.

Bafera, J., and S. Kleinert. 2022. "Signaling Theory in Entrepreneurship Research: A Systematic Review and Research Agenda." *Entrepreneurship Theory and Practice* 47, no. 6: 2419–2464. https://doi.org/10.1177/10422587221138489.

Bail, C. A. 2024. "Can Generative AI Improve Social Science?" *Proceedings of the National Academy of Sciences* 121, no. 21: e2314021121. https://doi.org/10.1073/pnas.2314021121.

Bankins, S., A. C. Ocampo, M. Marrone, S. L. D. Restubog, and S. E. Woo. 2024. "A Multilevel Review of Artificial Intelligence in Organizations: Implications for Organizational Behavior Research and Practice." *Journal of Organizational Behavior* 45, no. 2: 159–182. https://doi.org/10.1002/job.2735.

Bansal, G., T. Wu, J. Zhou, et al. 2021. "Does the Whole Exceed Its Parts? The Effect of AI Explanations on Complementary Team Performance."




10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

*CHI Conference on Human Factors in Computing Systems* 81: 1–16. https://doi.org/10.1145/3411764.3445717.

Bergh, D. D., B. L. Connelly, D. J. Ketchen, and L. M. Shannon. 2014. “Signalling Theory and Equilibrium in Strategic Management Research: An Assessment and a Research Agenda.” *Journal of Management Studies* 51, no. 8: 1334–1360. https://doi.org/10.1111/joms.12097.

Bigman, Y. E., and K. Gray. 2018. “People Are Averse to Machines Making Moral Decisions.” *Cognition* 181: 21–34. https://doi.org/10.1016/j.cognition.2018.08.003.

Bogert, E., A. Schecter, and R. Watson. 2021. “Humans Rely More on Algorithms Than Social Influence as a Task Becomes More Difficult.” *Scientific Reports* 11: 8028. https://doi.org/10.1038/s41598-021-87480-9.

Bolukbasi, T., K. Chang, J. Zou, V. Saligrama, and A. T. Kalai. 2016. “Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings.” *Proceedings of the 30th International Conference on Neural Information Processing Systems* 29: 4356–4364. https://doi.org/10.5555/3157382.3157584.

Branzei, O., T. J. Ursacki-Bryant, I. Vertinsky, and W. Zhang. 2004. “The Formation of Green Strategies in Chinese Firms: Matching Corporate Environmental Responses and Individual Principles.” *Strategic Management Journal* 25: 1075–1095. https://doi.org/10.1002/smj.409.

van den Broek, E., A. Sergeeva, and M. Huysman. 2021. “When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring.” *MIS Quarterly* 45, no. 3: 1557–1580. https://doi.org/10.25300/MISQ/2021/16559.

Brougham, D., and J. Haar. 2020. “Technological Disruption and Employment: The Influence on Job Insecurity and Turnover Intentions: A Multi-Country Study.” *Technological Forecasting and Social Change* 161: 120276. https://doi.org/10.1016/j.techfore.2020.120276.

Buçinca, Z., M. B. Malaya, and K. Z. Gajos. 2021. “To Trust or to Think.” *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW1: 1–21. https://doi.org/10.1145/3449287.

Bundorf, M. K., M. Polyakova, and M. Tai-Seale. 2019. “How Do Humans Interact With Algorithms? Experimental Evidence From Health Insurance.” National Bureau of Economic Research Working Papers 25976. https://doi.org/10.3386/w25976.

Buolamwini, J., and T. Gebru. 2018. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* 81: 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html.

Burton, J. W., M.-K. Stein, and T. B. Jensen. 2020. “A Systematic Review of Algorithm Aversion in Augmented Decision Making.” *Journal of Behavioral Decision Making* 33, no. 2: 220–239. https://doi.org/10.1002/bdm.2155.

Busenitz, L. W., J. O. Fiet, and D. D. Moesel. 2005. “Signaling in Venture Capitalist—New Venture Team Funding Decisions: Does It Indicate Long-Term Venture Outcomes?” *Entrepreneurship Theory and Practice* 29, no. 1: 1–12. https://doi.org/10.1111/j.1540-6520.2005.00066.x.

Çalışkan, A., J. J. Bryson, and A. Narayanan. 2017. “Semantics Derived Automatically From Language Corpora Contain Human-Like Biases.” *Science* 356, no. 6334: 183–186. https://doi.org/10.1126/science.aal4230.

Castelo, N. 2019. *Blurring the Line Between Human and Machine: Marketing Artificial Intelligence* [Doctoral Dissertation, Columbia University]. New York: Columbia University Libraries. https://doi.org/10.7916/d8-k7vk-0s40.

Chong, L., G. Zhang, K. Goucher-Lambert, K. Kotovsky, and J. Cagan. 2022. “Human Confidence in Artificial Intelligence and in Themselves: The Evolution and Impact of Confidence on Adoption of AI Advice.” *Computers in Human Behavior* 127: 107018. https://doi.org/10.1016/j.chb.2021.107018.

Chugunova, M., and D. Sele. 2022. “We and It: An Interdisciplinary Review of the Experimental Evidence on How Humans Interact With Machines.” *Journal of Behavioral and Experimental Economics* 99: 101897. https://doi.org/10.1016/j.socec.2022.101897.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates. https://doi.org/10.1016/C2013-0-10517-X.

Cohen, B. D., and T. J. Dean. 2005. “Information Asymmetry and Investor Valuation of IPOs: Top Management Team Legitimacy as a Capital Market Signal.” *Strategic Management Journal* 26, no. 7: 683–690. https://doi.org/10.1002/smj.463.

Connelly, B. L., S. T. Certo, R. D. Ireland, and C. R. Reutzel. 2011. “Signaling Theory: A Review and Assessment.” *Journal of Management* 37, no. 1: 39–67. https://doi.org/10.1177/0149206310388419.

Davila, A., G. Foster, and M. Gupta. 2003. “Venture Capital Financing and the Growth of Startup Firms.” *Journal of Business Venturing* 18, no. 6: 689–708. https://doi.org/10.1016/S0883-9026(02)00127-1.

Deci, E. L., and R. M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press.

Dellermann, D., P. Ebel, M. Söllner, and J. M. Leimeister. 2019. “Hybrid Intelligence.” *Business & Information Systems Engineering* 61, no. 5: 637–643. https://doi.org/10.1007/s12599-019-00595-2.

Dietvorst, B. J., and S. Bharti. 2020. “People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error.” *Psychological Science* 31, no. 10: 1302–1314. https://doi.org/10.1177/0956797620948841.

Dietvorst, B. J., J. P. Simmons, and C. Massey. 2015. “Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err.” *Journal of Experimental Psychology: General* 144, no. 1: 114–126. https://doi.org/10.1037/xge0000033.

Dietvorst, B. J., J. P. Simmons, and C. Massey. 2018. “Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms if They Can (Even Slightly) Modify Them.” *Management Science* 64, no. 3: 1155–1170. https://doi.org/10.1287/mnsc.2016.2643.

Drover, W., M. Wood, and A. C. Corbett. 2018. “Toward a Cognitive View of Signalling Theory: Individual Attention and Signal Set Interpretation.” *Journal of Management Studies* 55, no. 2: 209–231. https://doi.org/10.1111/joms.12282.

Đula, I., T. Berberena, K. Keplinger, and M. Wirzberger. 2023. “Hooked on Artificial Agents: A Systems Thinking Perspective.” *Frontiers in Behavioral Economics* 2. https://doi.org/10.3389/frbhe.2023.1223281.

Einola, K., and V. Khoreva. 2022. “Best Friend or Broken Tool? Exploring the Co-Existence of Humans and Artificial Intelligence in the Workplace Ecosystem.” *Human Resource Management* 62, no. 1: 117–135. https://doi.org/10.1002/hrm.22147.

Fumagalli, E., S. Rezaei, and A. Salomons. 2022. “OK Computer: Worker Perceptions of Algorithmic Recruitment.” *Research Policy* 51, no. 2: 104420. https://doi.org/10.1016/j.respol.2021.104420.

Gagné, M., S. K. Parker, M. A. Griffin, et al. 2022. “Understanding and Shaping the Future of Work With Self-Determination Theory.” *Nature Reviews Psychology* 1, no. 7: 378–392. https://doi.org/10.1038/s44159-022-00056-w.

Gillath, O., T. Ai, M. Branicky, S. Keshmiri, R. Davison, and R. Spaulding. 2021. “Attachment and Trust in Artificial Intelligence.” *Computers in Human Behavior* 115: 106607. https://doi.org/10.1016/j.chb.2020.106607.

Glikson, E., and A. W. Woolley. 2020. “Human Trust in Artificial Intelligence: Review of Empirical Research.” *Academy of Management Annals* 14, no. 2: 627–660. https://doi.org/10.5465/annals.2018.0057.

Gonzalez, M. F., W. Liu, L. Shirase, et al. 2022. “Allying With AI? Reactions Toward Human-Based, AI/ML-Based, and Augmented



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Hiring Processes." *Computers in Human Behavior* 130: 107–179. https://doi.org/10.1016/j.chb.2022.107179.

Götz, F. M., R. Maertens, S. Loomba, and S. van der Linden. 2023. "Let the Algorithm Speak: How to Use Neural Networks for Automatic Item Generation in Psychological Scale Development." *Psychological Methods, Forthcoming* 29: 494–518. https://doi.org/10.1037/met0000540.

Green, B., and Y. Chen. 2019. "Disparate Interactions: An Algorithm-In-The-Loop Analysis of Fairness in Risk Assessments." In *Proceedings of the Conference on Fairness, Accountability and Transparency*, 90–99. https://doi.org/10.1145/3287560.3287563.

Grother, P., M. Ngan, and K. Hanaoka. 2019. *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects* (Report No. NISTR 8280). Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8280.

Grundke, A. 2024. "If Machines Outperform Humans: Status Threat Evoked by and Willingness to Interact With Sophisticated Machines in a Work-Related Context." *Behaviour & Information Technology* 43, no. 7: 1348–1364. https://doi.org/10.1080/0144929x.2023.2210688.

Guest, D., K. Sanders, R. Rodrigues, and T. C. Oliveira. 2021. "Signalling Theory as a Framework for Analysing Human Resource Management Processes and Integrating Human Resource Attribution Theories: A Conceptual Analysis and Empirical Exploration." *Human Resource Management Journal* 31, no. 3: 796–818. https://doi.org/10.1111/1748-8583.12326.

Guo, G., and H. Zhao. 2000. "Multilevel Modeling for Binary Data." *Annual Review of Sociology* 26: 441–462. https://www.jstor.org/stable/223452.

Haesevoets, T., D. De Cremer, K. Dierckx, and A. Van Hiel. 2021. "Human-Machine Collaboration in Managerial Decision Making." *Computers in Human Behavior* 119: 106730. https://doi.org/10.1016/j.chb.2021.106730.

Himmelstein, M., and D. V. Budescu. 2023. "Preference for Human or Algorithmic Forecasting Advice Does Not Predict if and How It Is Used." *Journal of Behavioral Decision Making* 36, no. 1: e2285. https://doi.org/10.1002/bdm.2285.

Hollenbeck, J. R., B. Beersma, and M. E. Schouten. 2012. "Beyond Team Types and Taxonomies: A Dimensional Scaling Conceptualization for Team Description." *Academy of Management Review* 37: 82–106. https://doi.org/10.5465/amr.2010.0181.

Howard, J. J., L. R. Rabbitt, and Y. B. Sirotin. 2020. "Human-Algorithm Teaming in Face Recognition: How Algorithm Outcomes Cognitively Bias Human Decision-Making." *PLoS ONE* 15, no. 8: e0237855. https://doi.org/10.1371/journal.pone.0237855.

Hu, L., and P. M. Bentler. 1998. "Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification." *Psychological Methods* 3: 424–453. https://doi.org/10.1037/1082-989X.3.4.424.

Innocenti, S., and M. Golin. 2022. "Human Capital Investment and Perceived Automation Risks: Evidence From 16 Countries." *Journal of Economic Behavior and Organization* 195: 27–41. https://doi.org/10.1016/j.jebo.2021.12.027.

Jacobs, M., M. F. Pradier, T. H. McCoy, R. H. Perlis, F. Doshi-Velez, and K. Z. Gajos. 2021. "How Machine-Learning Recommendations Influence Clinician Treatment Selections: The Example of Antidepressant Selection." *Translational Psychiatry* 11, no. 1. https://doi.org/10.1038/s41398-021-01224-x.

Jago, A. S. 2019. "Algorithms and Authenticity." *Academy of Management Discoveries* 5, no. 1: 38–56. https://doi.org/10.5465/amd.2017.0002.

Jussupow, E., I. Benbasat, and A. Heinzl. 2020. "Why Are We Averse Towards Algorithms? A Comprehensive Literature Review on Algorithm Aversion." In *Proceedings of the 28th European Conference on Information Systems*. https://aisel.aisnet.org/ecis2020_rp/168.

Jussupow, E., I. Benbasat, and A. Heinzl. 2024. "An Integrative Perspective on Algorithm Aversion and Appreciation in Decision-Making." *MIS Quarterly* 48, no. 4: 1575–1590. https://doi.org/10.25300/MISQ/2024/18512.

Kaushik, M. 2022. "Artificial Intelligence (AI)." In *Intelligent System Algorithms and Applications in Science and Technology*, edited by S. Pathak, P. K. Bhatt, S. K. Singh, A. Tripathi, and P. K. Pandey, 119–133. New York: Apple Academic Press Inc. https://doi.org/10.1201/9781003187059.

Keding, C., and P. Meissner. 2021. "Managerial Overreliance on AI-Augmented Decision-Making Processes: How the Use of AI-Based Advisory Systems Shapes Choice Behavior in R&D Investment Decisions." *Technological Forecasting and Social Change* 171: 120970. https://doi.org/10.1016/j.techfore.2021.120970.

Kennedy, R., P. Waggoner, and M. Ward. 2021. "Trust in Public Policy Algorithms." *Journal of Politics* 84: 1132–1148. https://doi.org/10.1086/716283.

Kleinert, S., and K. Mochkabadi. 2022. "Gender Stereotypes in Equity Crowdfunding: The Effect of Gender Bias on the Interpretation of Quality Signals." *Journal of Technology Transfer* 47, no. 6: 1640–1661. https://doi.org/10.1007/s10961-021-09892-z.

Köchling, A., and M. C. Wehner. 2020. "Discriminated by an Algorithm: A Systematic Review of Discrimination and Fairness by Algorithmic Decision-Making in the Context of HR Recruitment and HR Development." *Business Research* 13, no. 3: 795–848. https://doi.org/10.1007/s40685-020-00134-w.

Lai, V., and C. Tan. 2019. "On Human Predictions With Explanations and Predictions of Machine Learning Models: A Case Study on Deception Detection." In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 29–38. https://doi.org/10.1145/3287560.3287590.

Leslie, L. M. 2019. "Diversity Initiative Effectiveness: A Typological Theory of Unintended Consequences." *Academy of Management Review* 44, no. 3: 538–563. https://doi.org/10.5465/amr.2017.0087.

Logg, J. M., J. A. Minson, and D. A. Moore. 2019. "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment." *Organizational Behavior and Human Decision Processes* 151: 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005.

Longoni, C., A. Bonezzi, and C. K. Morewedge. 2019. "Resistance to Medical Artificial Intelligence." *Journal of Consumer Research* 46, no. 4: 629–650. https://doi.org/10.1093/jcr/ucz013.

Lundberg, S., B. G. Nair, M. S. Vavilala, et al. 2018. "Explainable Machine-Learning Predictions for the Prevention of Hypoxaemia During Surgery." *Nature Biomedical Engineering* 2, no. 10: 749–760. https://doi.org/10.1038/s41551-018-0304-0.

Madhavan, P., D. A. Wiegmann, and F. C. Lacson. 2006. "Automation Failures on Tasks Easily Performed by Operators Undermine Trust in Automated Aids." *Human Factors* 48, no. 2: 241–256. https://doi.org/10.1518/001872006777724408.

Mahmud, H., A. K. M. N. Islam, S. I. Ahmed, and K. Smolander. 2022. "What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion." *Technological Forecasting and Social Change* 175: 121390. https://doi.org/10.1016/j.techfore.2021.121390.

Maier, S. B., E. Jussupow, and A. Heinzl. 2019. "Good, Bad, OR Both? Measurement of Physician's Ambivalent Attitudes Towards AI." In *Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden, June 8–14, 2019*. https://aisel.aisnet.org/ecis2019_rp/115.

Mantello, P., M. T. Ho, M. H. Nguyen, and Q. H. Vuong. 2023. "Bosses Without a Heart: Socio-Demographic and Cross-Cultural Determinants



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of Attitude Toward Emotional AI in the Workplace." *AI & SOCIETY* 38, no. 1: 97–119. https://doi.org/10.1016/j.ijinfomgt.2023.102623.

Marti, E., T. B. Lawrence, and C. W. Steele. 2024. "Constructing Envelopes: How Institutional Custodians Can Tame Disruptive Algorithms." *Academy of Management Journal, Forthcoming.* 67: 1273–1301. https://doi.org/10.5465/amj.2019.1343.

Mathieu, J. E., J. R. Hollenbeck, D. van Knippenberg, and D. R. Ilgen. 2017. "A Century of Work Teams in the Journal of Applied Psychology." *Journal of Applied Psychology* 102, no. 3: 452–467. https://doi.org/10.1037/apl0000128.

Matthews, M., S. Matthews, D. Wang, and T. K. Kelemen. 2022. "Tweet, Like, Subscribe! Understanding Leadership Through Social Media Use." *Leadership Quarterly* 33, no. 1: 101580. https://doi.org/10.1016/j.leaqua.2021.101580.

Memmert, L., and E. Bittner. 2022. "Complex Problem Solving Through Human-AI Collaboration: Literature Review on Research Contexts." In *Proceeding of the Hawaii International Conference on System Sciences*, 378–387. https://doi.org/10.24251/HICSS.2022.046.

Merritt, S. M. 2011. "Affective Processes in Human–Automation Interactions." *Human Factors* 53, no. 4: 356–370. https://doi.org/10.1177/0018720811411912.

Miller, T. 2019. "Explanation in Artificial Intelligence: Insights From the Social Sciences." *Artificial Intelligence* 267: 1–38. https://doi.org/10.1016/j.artint.2018.07.007.

Mosqueira-Rey, E., E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Á. Fernández-Leal. 2022. "Human-In-The-Loop Machine Learning: A State of the Art." *Artificial Intelligence Review* 56, no. 4: 3005–3054. https://doi.org/10.1007/s10462-022-10246-w.

Murray, A., J. Rhymer, and D. G. Sirmon. 2021. "Humans and Technology: Forms of Conjoined Agency in Organizations." *Academy of Management Review* 46, no. 3: 552–571. https://doi.org/10.5465/amr.2019.0186.

Newman, D. T., N. J. Fast, and D. Harmon. 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. https://doi.org/10.1016/j.obhdp.2020.03.008.

Noti, G., and Y. Chen. 2023. "Learning When to Advise Human Decision Makers." *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence* 339: 3038–3048. https://doi.org/10.24963/ijcai.2023/339.

Offenwanger, A., A. J. Milligan, M. Chang, J. Bullard, and D. Yoon. 2021. "Diagnosing Bias in the Gender Representation of HCI Research Participants: How It Happens and Where We Are." *CHI Conference on Human Factors in Computing Systems* 399: 1–18. https://doi.org/10.1145/3411764.3445383.

Park, H. D., and P. C. Patel. 2015. "How Does Ambiguity Influence IPO Underpricing? The Role of the Signalling Environment." *Journal of Management Studies* 52, no. 6: 796–818. https://doi.org/10.1111/joms.12132.

Patel, B. N., L. Rosenberg, G. Willcox, et al. 2019. "Human–Machine Partnership With Artificial Intelligence for Chest Radiograph Diagnosis." *npj Digital Medicine* 2, no. 1. https://doi.org/10.1038/s41746-019-0189-7.

Pethig, F., and J. Kroenung. 2022. "Biased Humans, (Un)biased Algorithms?" *Journal of Business Ethics* 183, no. 3: 637–652. https://doi.org/10.1007/s10551-022-05071-8.

Phillips, P. J., H. Wechsler, J. Huang, and P. J. Rauss. 1998. "The FERET Database and Evaluation Procedure for Face-Recognition Algorithms." *Image and Vision Computing* 16, no. 5: 295–306. https://doi.org/10.1016/s0262-8856(97)00070-x.

Plummer, L. A., T. H. Allison, and B. L. Connelly. 2016. "Better Together? Signaling Interactions in New Venture Pursuit of Initial External Capital." *Academy of Management Journal* 59, no. 5: 1585–1604. https://doi.org/10.5465/amj.2013.0100.

Raisch, S., and S. Krakowski. 2021. "Artificial Intelligence and Management: The Automation–Augmentation Paradox." *Academy of Management Review* 46, no. 1: 192–210. https://doi.org/10.5465/amr.2018.0072.

Russell, S. J., and P. Norvig. 2003. *Artificial Intelligence: A Modern Approach*. 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Saragih, M., and B. W. Morrison. 2022. "The Effect of Past Algorithmic Performance and Decision Significance on Algorithmic Advice Acceptance." *International Journal of Human Computer Interaction* 38, no. 13: 1228–1237. https://doi.org/10.1080/10447318.2021.1990518.

Seeber, I., L. Waizenegger, S. Seidel, S. Morana, I. Benbasat, and P. B. Lowry. 2020. "Collaborating With Technology-Based Autonomous Agents: Issues and Research Opportunities." *Internet Research* 30, no. 1: 1–18. https://doi.org/10.1108/INTR-12-2019-0503.

Sergeeva, E., A. Sergeeva, H. Tang, K. Bongard-Blanchy, and P. Szolovits. 2023. "Right, No Matter Why: AI Fact-Checking and AI Authority in Health-Related Inquiry Settings." arXiv:2310.14358. https://arxiv.org/abs/2310.14358.

Snijders, C., R. Conijn, E. de Fouw, and K. van Berlo. 2022. "Humans and Algorithms Detecting Fake News: Effects of Individual and Contextual Confidence on Trust in Algorithmic Advice." *International Journal of Human Computer Interaction* 39, no. 7: 1483–1494. https://doi.org/10.1080/10447318.2022.2097601.

Spence, M. 1973. "Job Market Signaling." *Quarterly Journal of Economics* 87, no. 3: 355. https://doi.org/10.2307/1882010.

StataCorp. 2019. Stata Statistical Software: Release 16. https://www.stata.com/.

Steigenberger, N., and H. Wilhelm. 2018. "Extending Signaling Theory to Rhetorical Signals: Evidence From Crowdfunding." *Organization Science* 29, no. 3: 529–546. https://doi.org/10.1287/orsc.2017.1195.

Suseno, Y., C. Chang, M. Hudík, and E. S. Fang. 2022. "Beliefs, Anxiety and Change Readiness for Artificial Intelligence Adoption Among Human Resource Managers: The Moderating Role of High-Performance Work Systems." *International Journal of Human Resource Management* 33, no. 6: 1209–1236. https://doi.org/10.1080/09585192.2021.1931408.

Swim, J. K., K. J. Aikin, W. S. Hall, and B. A. Hunter. 1995. "Sexism and Racism: Old-Fashioned and Modern Prejudices." *Journal of Personality and Social Psychology* 68, no. 2: 199–214. https://doi.org/10.1037/0022-3514.68.2.199.

Thurman, N., S. C. Lewis, and J. Kunert. 2019. "Algorithms, Automation, and News." *Digital Journalism* 7, no. 8: 980–992. https://doi.org/10.1080/21670811.2019.1685395.

Tong, S., N. Jia, X. Luo, and Z. Fang. 2021. "The Janus Face of Artificial Intelligence Feedback: Deployment Versus Disclosure Effects on Employee Performance." *Strategic Management Journal* 42, no. 9: 1600–1631. https://doi.org/10.1002/smj.3322.

Troth, A. C., P. J. Jordan, S. A. Lawrence, and H. H. Tse. 2012. "A Multilevel Model of Emotional Skills, Communication Performance, and Task Performance in Teams." *Journal of Organizational Behavior* 33, no. 5: 700–722. https://doi.org/10.1002/job.785.

Uhlmann, E. L., V. L. Brescoll, and É. Machery. 2010. "The Motives Underlying Stereotype-Based Discrimination Against Members of Stigmatized Groups." *Social Justice Research* 23, no. 1: 1–16. https://doi.org/10.1007/s11211-010-0110-7.

Vaccaro, M., A. Almaatouq, and T. Malone. 2024. "When Combinations of Humans and AI Are Useful: A Systematic Review and Meta-Analysis." *Nature Human Behaviour* 8: 2293–2303. https://doi.org/10.1038/s41562-024-02024-1.



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Vanacker, T., D. Forbes, M. Knockaert, and S. Manigart. 2020. "Signal Strength, Media Attention, and Resource Mobilization: Evidence From New Private Equity Firms." *Academy of Management Journal* 63, no. 4: 1082–1105. https://doi.org/10.5465/amj.2018.0356.

Walter, B., D. Kremmel, and B. Jäger. 2021. "The Impact of Lay Beliefs About AI on Adoption of Algorithmic Advice." *Marketing Letters* 33: 143–155. https://doi.org/10.1007/s11002-021-09589-1.

Wang, D., P. Maes, X. Ren, B. Shneiderman, Y. Shi, and Q. Wang. 2021. "Designing AI to Work WITH or FOR People?" In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*, vol. 154, 1–5. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3411763.3450394.

Wang, H., Y. Zhang, and M. Wan. 2021. "Linking High-Performance Work Systems and Employee Well-Being: A Multilevel Examination of the Roles of Organisation-Based Self-Esteem and Departmental Formalisation." *Human Resource Management Journal* 32, no. 1: 92–116. https://doi.org/10.1111/1748-8583.12391.

Xu, J., I. Benbasat, and R. T. Cenfetelli. 2020. "The Relative Effect of the Convergence of Product Recommendations From Various Online Sources." *Journal of Management Information Systems* 37, no. 3: 788–819. https://doi.org/10.1080/07421222.2020.1790192.

Yeomans, M., A. Shah, S. Mullainathan, and J. Kleinberg. 2019. "Making Sense of Recommendations." *Journal of Behavioral Decision Making* 32: 403–414. https://doi.org/10.1002/bdm.2118.

You, S., C. Yang, and X. Li. 2022. "Algorithmic Versus Human Advice: Does Presenting Prediction Performance Matter for Algorithm Appreciation?" *Journal of Management Information Systems* 39: 336–365. https://doi.org/10.1080/07421222.2022.2063553.

Zhang, Y., Q. V. Liao, and R. K. E. Bellamy. 2020. "Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 295–305. https://doi.org/10.1145/3351095.3372852.

Zhao, M., R. Simmons, and H. Admoni. 2022. "The Role of Adaptation in Collective Human-AI Teaming." *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12633.



10991379, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/job.2856 by MPI 360 Intelligent Systems, Wiley Online Library on [21/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License