Technovation 131 (2024) 102953

Contents lists available at ScienceDirect

Technovation

journal homepage: www.elsevier.com/locate/technovation

# Toward human-centered AI management: Methodological challenges and future directions

Mengchen Dong [a,*], Jean-François Bonnefon [b,c], Iyad Rahwan [a]

[a] Center for Humans and Machines, Max Planck Institute for Human Development, Germany
[b] Institute for Advanced Study in Toulouse, University of Toulouse 1 Capitole, Toulouse, France
[c] Toulouse School of Economics (TSM-R), University of Toulouse 1 Capitole, Toulouse, France

ARTICLE INFO

Handling Editor: Dr. Stelvia Matos

*Keywords:*
Artificial intelligence
Algorithmic management
Algorithm aversion
Algorithm appreciation
Future of work
Work design
Crowdsourcing

ABSTRACT

As algorithms powered by Artificial Intelligence (AI) are increasingly involved in the management of organizations, it becomes imperative to conduct human-centered AI management research and understand people's feelings and behaviors when machines gain power over humans. The two mainstream methods – vignette studies and case studies – reveal important but inconsistent insights. Here we discuss the respective limitations of vignette studies (affective forecasting errors, biased media coverage, and question substitution) and case studies (social desirability biases and lack of random assignment and control conditions), which may lead them to overrate negative and positive reactions to AI management, respectively. We further discuss the advantages of a third method for mitigating these limitations: field experiments on crowdsourced marketplaces. A proof-of-concept study on Amazon Mechanical Turk (Mturk; as a world-leading crowdsourcing platform) showed unique human reactions to AI management, which were not perfectly aligned with those in vignette or case studies. Participants ($N = 504$) did not differ significantly under AI versus human management, in terms of performance, intrinsic motivation, fairness perception, and commitment. We suggest that crowdsourced marketplaces can go beyond human research subject pools and become models of AI-managed workplaces, facilitating timely behavioral research and robust predictions on human-centered work designs and organizations.

## 1. Introduction

Algorithms powered by Artificial Intelligence (AI) are increasingly involved in the management of organizations, a development that has spurred much research oriented toward efficiency, revenue, and innovation (Frank et al., 2019; Frey and Osborne, 2017; Kellogg et al., 2020; MIT Work of the Future, 2019; Wood, 2021). The use of AI in managing workers has also become a frontier for understanding people's feelings and behaviors where machines gain power over humans (Cao et al., 2021; Curchod et al., 2020; Glikson and Woolley, 2020; Höddinghaus et al., 2021; Kellogg et al., 2020; Rahman, 2021; Wesche and Sonderegger, 2019). The two most common methodological paradigms to collect data on human reactions to AI management are *vignette studies* and *case studies*. In vignette studies, participants are presented with hypothetical scenarios in which they are managed by AI, and asked to anticipate how they would feel and behave. Case studies, on the other hand, recruit participants who work in a company that has already deployed AI management, and integrate data from surveys, interviews, text analysis, or observation, to gauge how they actually feel and behave in their dealings with AI managers.

Vignette studies and case studies can have different focuses. For example, vignette studies often focus on futuristic scenarios where AI management has not yet happened but can be foreseen to produce certain outcomes, whereas case studies focus mainly on the status quo where AI management has already been deployed and systematically changed how some people work. Relatedly, vignette studies and case studies can focus on different industries and occupations that employ people with different characteristics. As compared to vignette studies that can examine a wide range of domains and populations, case studies often concern jobs that are most susceptible to automation and AI replacement (e.g., comprising mechanical, more than analytical, intuitive or empathetic tasks; Frey and Osborne, 2017; Huang and Rust, 2018). And people who perform these tasks tend to be less skilled, less educated, and have a relatively disadvantaged socio-economic background (Frank et al., 2019; Frey and Osborne, 2017; Makarius et al., 2020).

* Corresponding author. Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany.
*E-mail address:* dong@mpib-berlin.mpg.de (M. Dong).

https://doi.org/10.1016/j.technovation.2024.102953
Received 10 May 2023; Received in revised form 21 November 2023; Accepted 6 January 2024
Available online 16 January 2024

Despite various differences, vignette studies and case studies both contribute important insights into human-centered AI management by investigating people's positive or negative attitudes toward AI management. Here we show that findings from these two sources of data are often inconsistent. Typically, participants in case studies display more positive attitudes toward AI managers than participants in vignette studies. We argue that the two methods reveal conflicting results partly because they each miss a crucial component. Case studies do not have experimental controls – which should allow us to compare people's feelings and behaviors under AI management with a clear baseline. Instead, vignette studies have experimental controls but are often not grounded in any actual experience; expected feelings and intended behaviors can be shaped more by idiosyncratic thinking about dystopian scenarios rather than actual experience.

We then propose *field experiments* as a method that can overcome the limitations of vignette studies (specifically, affective forecasting errors, biases linked to media coverage, and question substitution) and the limitations of case studies (specifically, lack of random assignment, lack of experimental controls, and potential insincerity of respondents due to job insecurity; see Fig. 1 for a summary). Crowdsourced marketplaces such as Amazon Mechanical Turk (Mturk) can provide an appropriate setting for these field experiments, using their original function as online labor markets (Horton et al., 2011). To replicate previous findings or support our reasoning, we conducted two empirical studies – one representative-sample survey on UK citizens ($N$ = 488) and one field experiment with Mturk workers ($N$ = 504). All the study materials, datasets and codebook, and analysis scripts can be found at https://tinyurl.com/AImanagement. In the main text, we summarize the key results of the representative-sample survey, and present the detailed methods and results of the field experiment on Mturk. The detailed method and other ancillary results of the survey study can be found in the Supplementary Materials (SM).

## 2. Literature review

### 2.1. Inconsistent findings between vignette and case studies

In order to highlight the central discrepancy between vignette and case studies, we focus on the dependent variables that they most commonly share, that is, measures of attitudes toward AI management. In vignette studies, attitudes are often negative, in line with studies on algorithm aversion in many other domains (e.g., medical, aesthetic, judicial; Bigman and Gray, 2018; Castelo et al., 2019; Köbis and Mossink, 2021; Longoni et al., 2019).

The aversion to AI managers in vignette studies is especially pronounced for the managerial roles that people deem to require psychological rather than analytical qualities (Castelo et al., 2019; Gonzalez et al., 2022; Lee, 2018; Morewedge, 2022; Ranganathan and Benson, 2020). Since people perceive AI to be successful at analytical and problem-solving tasks, they are relatively open to letting AI help them schedule and monitor their workflows (Lee, 2018; Logg et al., 2019; Raveendhran and Fast, 2021). In contrast, people perceive AI to be less successful at recognizing emotions and understanding natural language, which makes it bad at roles that require understanding and responding to feelings (Acikgoz et al., 2020; Bigman and Gray, 2018; Gray et al., 2007; Höddinghaus et al., 2021; Longoni et al., 2019). As a result, people often express strong aversions to seeing AI in managerial roles that require these psychological qualities, especially when the role implies evaluating workers and making decisions about their careers, a process that people imagine as unfair, dehumanizing, and demotivating (Acikgoz et al., 2020; Castelo et al., 2019; Gonzalez et al., 2022; Lee, 2018; Newman et al., 2020).

While participants in vignette studies typically express negative attitudes toward AI management, case studies often tell a different story. Here we list a few examples of this discrepancy. First, in vignette studies, participants imagine that they would be reluctant to share their personal feelings with AI managers (Acikgoz et al., 2020; Höddinghaus et al., 2021; see also in our representative survey) – but in case studies, workers can be quite comfortable with disclosing emotions to a computer (Lucas et al., 2014; von der Pütten et al., 2010). Second, in vignette studies, participants expect AI management to deprive them of autonomy and intrinsic motivation (Gonzalez et al., 2022; Raveendhran and Fast, 2021) – but in case studies, workers sometimes do not experience a lack of autonomy and can instead feel empowered by AI managers (Kusk and Bossen, 2022; Mohlmann and Zalmanson, 2018). Third, in vignette studies, participants do not believe AI managers could properly evaluate them (Acikgoz et al., 2020; Lee, 2018; Raveendhran and Fast, 2021) – but in case studies, workers are willing to cede control authority to AI leaders and trust its procedural fairness (Cormier et al., 2013; Gombolay et al., 2015; Lix and Valentine, 2020). Fourth, in vignette studies, people think that being evaluated by AI would impair their commitment to their job and the organization (Gonzalez et al., 2022; Raveendhran and Fast, 2021) – but in case studies, workers can actually achieve higher productivity when managed by AI (Bai et al., 2020; Kawaguchi, 2020; Lee et al., 2021; Ranganathan and Benson, 2020; Sun et al., 2021).

Positive reactions to AI management in case studies emerge in particular when work design features human-centered standards (Parker and Grote, 2022), for example, when AI managers make fair decisions (Kusk and Bossen, 2022; Lee et al., 2021; Mohlmann and Zalmanson, 2018), can integrate workers' own experience (Kawaguchi, 2020; Lee et al., 2021) or other superiors' oversight (Gonzalez et al., 2022; Kusk and Bossen, 2022). However, they do not alleviate ethical concerns related to objective features of AI management, including, for example, lack of explainability and accountability (Glikson and Woolley, 2020;

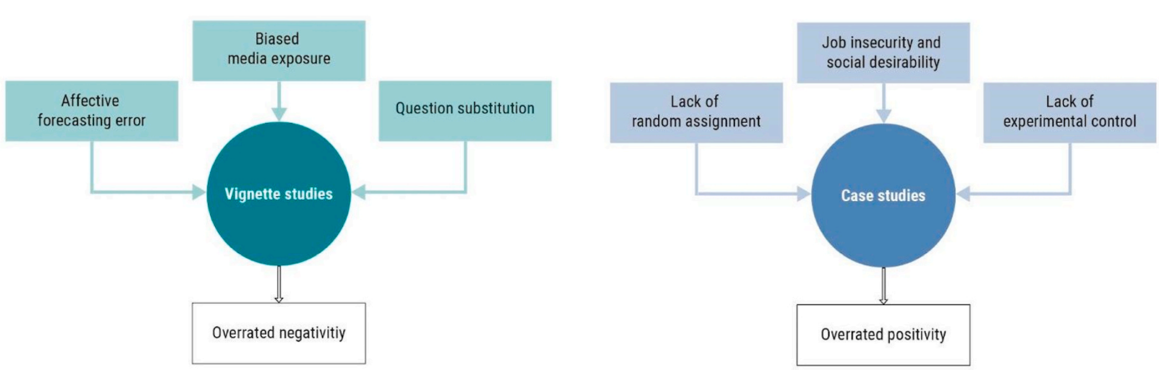


**Fig. 1.** A methodological summary of studies on human-centered AI management.
*Note.* The limitations of vignette studies and case studies can lead to overestimations of negative or positive reactions to AI management, respectively. These limitations also point to field experiments as a future direction, where crowdsourced labor markets can play an important role in simulating the future workplace and eliciting realistic reactions to AI management.

Kellogg et al., 2020), as well as information and power asymmetries (Crawford, 2021; Curchod et al., 2020; Kellogg et al., 2020; Rahman, 2021).

These and other findings suggest a gap between the attitudes people have toward AI management when they participate in vignette studies, and the attitudes workers actually have toward their AI managers in case studies. But how should we interpret this gap? In the next section, we suggest that vignette studies and case studies have different strengths and limitations, which may have contributed to different conclusions regarding people's attitudes toward AI managers.

### 2.2. Vignette and case studies have different limitations

The gap between vignette and case studies does not necessarily mean that one is right and the other is wrong. Instead, each method has its own limitations (and can complement each other in some respects). Vignette studies ask participants to imagine how they would feel about AI management, without having actually experienced it, which raises at least threats to external validity. First, people are usually poor judges of their reactions to future events: they underestimate their abilities to adapt to changes and overestimate the intensity and duration of their future emotional states (Wilson and Gilbert, 2005). This is likely to pose problems when people imagine being managed by AI, a situation that evokes a subordinate relation with strong status asymmetry and is known to trigger vigilant reactions (Kraus et al., 2011; Rucker et al., 2018). Indeed, research on human hierarchies has shown that when people imagine themselves in a low-status position, they feel less autonomy and control, are more vigilant toward threats in their social environment, and expect more hostile intentions from others in ambiguous situations (Kraus et al., 2011; Lount and Pettit, 2012; Rucker et al., 2018).

Second, the responses of participants to vignette studies, who do not have an actual experience of AI managers, are more likely to be shaped by the media coverage of AI management (Gonzalez et al., 2022; Haggadone et al., 2021) – and this media coverage is largely negative, as illustrated for example by news stories of the AI managers in Amazon warehouses, which are often depicted as exploitative and dehumanizing (Crawford, 2021; Dzieza, 2020; Soper, 2021).

Third, people may substitute difficult questions about how they might feel (once having experienced AI management) with easier ones like how they feel now (about switching to AI management; Kahneman and Frederick, 2002). Based on abstract descriptions and little experience, people may have a hard time anticipating their specific feelings (about fairness, autonomy, status, etc.). As a consequence, substitution can take place where people express preferences for the status quo (i.e., human management) and resistance to changes (i.e., AI management; Jachimowicz et al., 2019). The choice of human versus AI management, though, is not one especially realistic, since this decision is usually made for workers (Crawford, 2021; Rahman, 2021). When AI management is actually deployed, workers may express boredom and dissatisfaction with the tasks they receive, but they do not have a choice to switch back to human management (Bucher et al., 2021; Cormier et al., 2013).

Case studies do not have the problems above, since they collect measures from workers who actually have the experience of being managed by AI. As a result, these workers can respond based on their lived experience (Gonzalez et al., 2022; Haggadone et al., 2021; Makarius et al., 2020), which attenuates issues of affective forecasting, biased media exposure, and question substitution. Case studies, however, have their own limitations. As the science fiction writer William Gibson famously wrote, "The future is already here – it's just not very evenly distributed" (Gibson, 1999). While AI management may be in the cards for most jobs, it's not evenly distributed yet, which creates an endogeneity problem for case studies. By investigating organizations that have already deployed AI management, case studies automatically select sectors that are the most amenable to this management, and workers who chose to enroll or remain in these sectors, whose profile may thus be different from the general population. For example, workers currently under AI management often have lower skills and incomes (Frank et al., 2019; Frey and Osborne, 2017; Makarius et al., 2020), and their attitudes can partly result from their specific socioeconomic characteristics (Kraus et al., 2011; Lount and Pettit, 2012; Rucker et al., 2018).

Second, the validity of findings in case studies may be jeopardized by workers' feelings of job insecurity. When they are aware that their behavior is being observed and analyzed, people tend to behave in socially desirable ways. Workers who are managed by AI and feel that their job is not secure (which can be particularly true for workers with lower skills and incomes) may exaggerate their satisfaction with their AI managers because they believe (rightly or wrongly) that their answers are monitored and that they should behave in the way their hierarchy deems desirable (Arnold et al., 1985; Dunning et al., 2004).

Third, case studies on AI management often do not incorporate rigorous contrasts with control conditions such as human management, which can hinder inferences about whether human management strategies and work designs should differ from AI management ones (Kellogg et al., 2020; Parker and Grote, 2022). While it is easy for vignette studies to provide participants with a controlled counterfactual, and to provide quantitative evidence of how differently they feel about AI management compared to the human management counterfactual, it is challenging to do the same in case studies, as it is typically impossible to intervene on workers' actual organization and implement experimental manipulations. There are exceptions; for example, Lix and Valentine (2020) examined how freelancers react to platform rewards differently before versus after removing an AI-based scoring system. Ranganathan and Benson (2020) investigated work motivation and productivity after implementing AI versus human management on two existing production lines. Despite their valuable insights into the differences between human and AI management in the field, these quasi-experiments can also be biased by the non-randomized sequence of different treatments or other confounding factors in quasi-experimental groups.

To sum up, vignette studies can suffer from affective forecasting, biased media exposure, and question substitution, and case studies, which do not have these problems, can still suffer from endogeneity, social desirability biases, and lack of control conditions – problems that vignette studies do not face to the same extent. Ideally, research would benefit from a method that combines the respective strengths of vignette and case studies, and mitigate their respective threats to validity. We now argue that field experiments can be a practical and effective solution, especially if they draw on online labor markets that behavioral scientists already have experience with (Buhrmester et al., 2011; Paolacci et al., 2010).

### 2.3. Field experiments in crowdsourced marketplaces

Vignette and case studies of AI management often deliver inconsistent results, and the gap between these results may be due to their respective limitations. To make progress toward a resolution, researchers would need to conduct properly controlled experiments in the field, that is, in a setting that is as close as possible to real employment while allowing random assignment to treatment and control conditions, and allowing research participants to express sincere attitudes toward AI management, without fear about personal consequences. This is a very high bar; to facilitate more timely research on human-centered AI management (White et al., 2022), we suggest that there is an employment context that many researchers are already familiar with, which can provide an acceptable approximation: crowdsourced marketplaces, of which the best-known example is Amazon Mechanical Turk (Mturk). Crowdsourced marketplaces have become an everyday tool for many social and behavioral scientists, following seminal research that validated them as a data collection platform for social science studies (Buhrmester et al., 2011; Horton et al., 2011). Within five years, social science journals with an impact factor greater than 2.5 published 10 times more papers that recruited participants from Mturk, which increased from fewer than 50 in 2011 to more than 500 in 2015 (Chandler and Shapiro, 2016). Management research witnessed a similar trend using data from Mturk, with 6 papers doing so in 2012 to 133 in 2019 (Aguinis et al., 2021). Social scientists have run so many surveys and experiments on Mturk that many have come to see it as a global, extended department subject pool (Buhrmester et al., 2011; Horton et al., 2011; Paolacci et al., 2010), overlooking the fact that it is primarily an online labor market, where workers build business connections with clients, with flexible hours and a substantial level of autonomy.

Crowdsourced marketplaces may not only be regarded as an approximation to other employment contexts; they are by nature growing economic institutions and increasingly share characteristics with other workplaces in the post-pandemic era (Parker and Grote, 2022). Triggered by the pandemic and enabled by digital technologies, a larger proportion of the workforce across various industries has experienced a digital transformation. By working from home, people can save commuting time and have flexible schedules. Through virtual work activities, however, they may also experience loneliness and blurred work-life boundaries (Wood et al., 2022). These shared characteristics establish a common ground for crowdsourced and other working contexts.

It should be noted, however, that crowdsourced marketplaces can face a generalizability problem when it comes to highly heterogeneous work contexts or worker characteristics. Even though crowdsourced marketplaces are a good example of economic institutions that facilitate micro-work, crowdsourcing, and the so-called gig economy, they lack the long-term commitments and physical interactions that still predominate in many employment situations. For example, the way people experience performance monitoring, job autonomy, psychological contract, and social support systematically differ between virtual versus physical work contexts (Kellogg et al., 2020; Wood, 2021; Wood et al., 2022), and the way these factors influence people's attitude toward AI management may also differ. Moreover, workers on crowdsourced marketplaces have demographic characteristics that may also influence their attitudes toward AI management. For example, Mturk workers are mainly based in the United States; as compared to the general population of US working adults, Mturk workers are generally representative on gender, ethnicity, and income, but are relatively younger and better educated (e.g., more than 50% of them have a college degree, compared to 36% of working US adults over 18 years old; Geiger, 2016; Moss, 2020). Thus, findings based on these samples may not generalize to other cultural contexts, or work groups featuring an older age or lower level of education. Nonetheless, in the context of AI management, we believe that crowdsourced marketplaces can at least mitigate the limitations of vignette and case studies and help consolidate previous findings on human-centered AI management.

Conducting field experiments on AI management in crowdsourced marketplaces can overcome some limitations of vignette studies. Most importantly, these field experiments make it possible to hire participants to complete compensated tasks under the scrutiny of a real AI manager. As a result, participants do not have to imagine how they would feel or behave in this context – they are directly experiencing the situation, under incentive-compatible conditions. Just as important, these experiments do not require deception, which may induce biases in interpreting human reactions to AI managers (Tong et al., 2021). Crowdsourced marketplaces like Mturk provide simple Application Programming Interfaces (APIs) and flexible integration with experimental design tools (e.g., Qualtrics, oTree) that allow researchers to pre-program algorithms to allocate tasks to participants, monitor their performance, and determine their rewards and subsequent tasks.

Conducting field experiments on AI management in crowdsourced marketplaces can also overcome some limitations of case studies. First, these field experiments make it easier to recruit diverse samples of workers (Aguinis et al., 2021; Buhrmester et al., 2011), alleviating the endogeneity problem faced by case studies. For example, Mturk system allows to select relatively skilled or less skilled workers (based on their number of completed tasks and approval rate) and workers based in different U.S. states or other countries. Using pre-screening surveys, recruiters can implement customized questions and standards, and invite workers who provide targeted answers (e.g., a particular education level) to join the focal tasks later linked through their unique Worker ID. A diverse composition of workers can then be achieved by assigning a balanced quota for these system or customized qualifications (e.g., 50% below and 50% above college education). Second, these field experiments make it possible to run tightly controlled treatments and to randomly assign workers to these treatments. Third, participants in field experiments can express themselves without fear of negative consequences. If they are fully informed of the way their compensation is calculated, they can trust that their feedback about their managers will not impact their payoff. Furthermore, the short-term and anonymous nature of their employment means that they can express themselves freely, without consequences for their future jobs within the crowdsourced marketplace.

## 3. Summary of a representative survey

We first conducted a survey study, to synthesize previous findings on people's reactions to AI management. This was done since, as we reasoned earlier, vignette and case studies can have different focuses in terms of, for example, industries (relatedly, managerial functions) and worker characteristics. Relatedly, vignette and case studies also suggest different reasons why people approve or disapprove of AI management. To provide an overview, we surveyed 488 UK citizens through a panel company Respondi. We aimed for a sample that was representative on age ($M_{\text{age}} = 47.05$, $SD = 16.09$, with 6 missing), gender (55.3% male, 42.2% female, and 2.5% other), and education level (16.6% lower secondary, 30.1% upper secondary, 16.8% post-secondary, and 18.0% advanced-level tertiary education), and answered questions about (A) their general attitudes toward AI managers, as well as AI assistants and co-workers, (B) their acceptance of AI serving different managerial functions, and (C) potential reasons why they approve or disapprove of AI management. The main findings are summarized in Fig. 2.

In general, 66% of our respondents disapproved of the replacement of human managers with AI managers, which was twice the rate at which they disapproved of the use of AI in non-managerial roles such as assistants or co-workers (34.5% and 29.6%, respectively; see Fig. 2A). The aversion to AI management was extremely strong when AI management was deployed to understand personal feelings; 73% of respondents in our survey expressed such aversion (see Fig. 2B). Among various reasons why people might disapprove of AI management, as shown in Fig. 2C, autonomy and fairness concerns were strongly negatively correlated with people's acceptance of AI managers (e.g., "AI managers would deprive my autonomy", $r = -0.46$; "AI managers would leave me with no channel to appeal", $r = -0.43$; or "AI managers would disrespect me", $r = -0.40$) were strongly negatively correlated with people's acceptance of AI managers, more so than AI managers' competencies ("AI managers would be competent to do their job"; $r = 0.15$).

## 4. A proof-of-concept field experiment on Mturk

We reasoned that field experiments can mitigate some limitations of vignette studies (which may overestimate negative attitudes toward AI managers) and case studies (which may overestimate positive attitudes toward AI managers), and better simulate the future of AI management. As an illustration of what a field experiment on AI management in a crowdsourced marketplace might reveal, we conducted a proof-of-concept experiment on Mturk, focusing on actual performance and psychological reactions such as autonomy, fairness, intrinsic motivation, and commitment – as suggested by our representative survey and other previous studies (Acikgoz et al., 2020; Gonzalez et al., 2022; Lee, 2018;

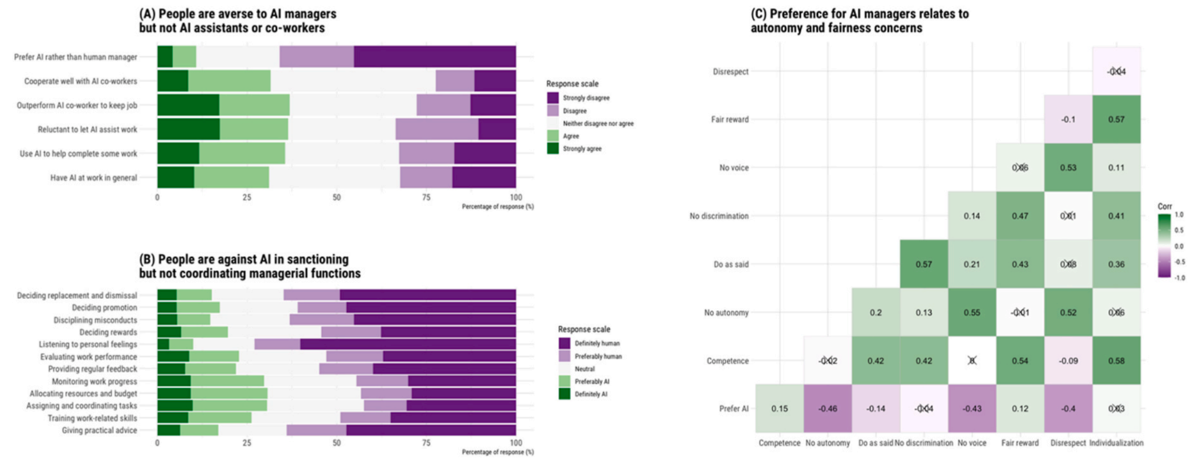


**Fig. 2.** Summary of the representative-sample survey on AI management (N = 488).

Mohlmann and Zalmanson, 2018; Newman et al., 2020).

We assigned Mturk workers to be managed either by a human or by an algorithm. Workers had to complete a personality test, and knew that their bonus pay would depend on whether their manager (which they knew to be either a human or an algorithm) judged them to have completed this task seriously. We know from previous research that people believe this management task to require high psychological ability, and that they express doubts about the capacity of AI managers to perform it adequately (Castelo et al., 2019). The study was pre-registered on Open Science Framework at https://tinyurl.com/PreregisterAImanagement.

### 4.1. Method

#### 4.1.1. Participants

As specified in our pre-registration, we recruited 504 participants (298 males; $M_{\text{age}} = 39.6$, $SD = 11.7$) on Mturk to complete a study about Personality Test Evaluation on two consecutive days. Participants completed the personality test on Day 1, and were contacted through their Mturk worker ID to receive the evaluation results on the following Day 2. And 363 participants (72.0%; 210 males; $M_{\text{age}} = 39.7$, $SD = 11.9$) responded and returned to complete the experiment on Day 2.

#### 4.1.2. Design

We employed a 2 (manager: algorithm vs. human) by 2 (evaluation: correct or incorrect) between-participants design. Participants were randomly assigned to either an algorithmic manager ($n = 266$) or a human manager ($n = 238$) condition on Day 1. They then received either a correct ($n = 181$) or an incorrect ($n = 182$) evaluation result on Day 2.

#### 4.1.3. Procedure

We introduced the manager conditions as "whether you treat the personality test seriously will either be assessed by (1) an algorithm embedded in this survey or (2) a researcher involved in this study". Participants were then randomly assigned to one of the two manager conditions, being asked about the manager's eligibility to evaluate their seriousness ("How ineligible or eligible do you think the algorithm/researcher is in evaluating whether you treat the personality test seriously?"; from 1 = *definitely ineligible* to 7 = *definitely eligible*), and then started the 100-item HEXACO personality test (Lee and Ashton, 2018), which took about 15 min. At the end of the personality test, participants answered four questions about intrinsic motivation (Deci et al., 1994; including effort ["I put a lot of effort into the personality test"], enjoyment ["I enjoyed doing the personality test very much"], nervosity ["I did not feel nervous at all while doing the personality test"], and autonomy ["I believe I had some choice about doing the personality test"]; from 1 = *strongly disagree* to 7 = *strongly agree*), and self-reported their seriousness ("Do you think you were unserious or serious about completing the personality test?"; from 1 = *extremely unserious* to 7 = *extremely serious*) and estimated seriousness evaluation ("Do you think the algorithm/researcher will evaluate you as unserious or serious about completing the personality test?"; from 1 = *extremely unserious* to 7 = *extremely serious*) from the manager.

After one day, participants in the Day-1 study received an invitation to take part in the Day-2 study. They were first reminded that "it was an algorithm embedded in the survey [in the algorithmic manager condition]/a researcher involved in this study [in the human manager condition] that evaluated whether you treated the personality test seriously and whether to give you the $1.3 bonus for the personality test". Participants were then informed that "the algorithm/researcher evaluated you as SERIOUS/UNSERIOUS about completing the personality test. Therefore, you will/will not receive the $1.3 bonus for the personality test". Identical validated criteria[1] were adopted to distinguish serious from unserious workers (Barends and De Vries, 2019), either by a researcher in the human manager condition, or by actual algorithms (i.e., Java scripts) embedded within the task in the algorithmic manager condition. In the incorrect evaluation condition, serious participants received an unserious evaluation without the bonus pay, while unserious participants received a serious evaluation with the bonus pay. Instead, in the correct evaluation condition, serious participants were rated as serious and unserious participants as unserious. After receiving the evaluation results, participants indicated their perceptions of procedural (e.g., "The evaluation procedure is free of bias"; $\alpha = 0.91$ for four items) and distributive (e.g., "The evaluation result reflects the effort I put into the work"; $\alpha = 0.93$ for three items) fairness (rated on a 7-point scale from 1 = *strongly disagree* to 7 = *strongly agree*; Colquitt, 2001). Participants each answered two questions about their intentions to enroll in a similar task in the future, either managed by a human or an algorithm ("I am willing to take part in a similar personality test if my seriousness about completing the test is evaluated by a researcher/an

[1] According to Barends and De Vries (2019), we applied three criteria to classify serious and unserious workers. Participants who meet any of the three exclusion criteria are flagged as unserious.(1) After averaging all pre-recoded HEXACO items, participants whose standard deviation is below 0.70 are identified as unserious. (2) After reverse coding the negatively keyed items and calculating the standard deviations of six HEXACO dimensions, participants whose averaged standard deviation is above 1.60 are identified as unserious. (3) After reverse coding the third infrequency item (i.e., "I can count from one to twenty-five"), "strongly disagree" and "disagree" are coded as 1 and other choices are coded as 0. Participants whose averaged score of four infrequency items is below 0.75 are identified as unserious.

algorithm"; from 1 = *strongly disagree* to 7 = *strongly agree*).

Participants were debriefed at the end of the Day-2 study. They were informed about their actual evaluation result (i.e., as serious or unserious). We sent bonuses to all participants who were actually serious or received incorrect evaluations about them being serious. In other words, only participants who were unserious and received correct evaluations about them being unserious did not receive the $1.3 bonus.

### 4.2. Results

Following the pre-registration, we analyzed Mturk workers' actual performance (i.e., being serious or unserious), intrinsic motivation (including autonomy), fairness perceptions, and intended future commitment (as summarized in Fig. 3). We performed the analyses following both the Frequentist and Bayesian approaches. For the latter approach, we reported the Bayes Factor in favor of the null hypothesis over the alternative ($BF_{01}$).

#### 4.2.1. Actual performance

With 323 out of the 504 workers (64.1%) being serious about the task, neither behavioral (63.9% vs. 64.3%; $z = 0.10$, $p = 0.92$; $BF_{01} = 25.64$) nor self-reported ($M = 6.37$, $SD = 0.81$ vs. $M = 6.36$, $SD = 0.77$; $t = 0.16$, $p = 0.87$; $BF_{01} = 55.56$) seriousness differed in the algorithmic versus the human manager condition. The 363 participants (72.0%) who returned to complete the Day-2 study were also equivalently distributed in the two manager conditions (73.7% vs. 70.2%; $z = 0.88$, $p = 0.38$; $BF_{01} = 17.54$).

#### 4.2.2. Intrinsic motivation

Neither the overall motivation (multivariate $F$ (4, 499) = 0.49, $p = 0.75$) nor specific reports of effort, enjoyment, nervosity, or autonomy ($p > 0.18$; $BF_{01} > 22.73$) differed in the algorithmic versus the human manager condition (see Fig. 3A).

#### 4.2.3. Fairness perceptions

As shown in Fig. 3B, for both serious and unserious workers, perceived procedural and distributive fairness were contingent on the


**Fig. 3.** Summary of the Mturk field experiment on AI management (N = 504).

valence of evaluation (multivariate $F$ (2, 216) = 45.12, $p$ < 0.001 for serious workers, and multivariate $F$ (2, 137) = 45.12, $p$ = 0.005 for unserious workers) but not manager condition (multivariate $F$ (2, 216) = 1.22, $p$ = 0.30 for serious workers, and multivariate $F$ (2, 137) = 2.77, $p$ = 0.07 for unserious workers). Negative (vs. positive) evaluation and rejected bonus pay meant unfair (vs. fair) treatment for serious workers. They did not see unfair (vs. fair) treatment by an algorithmic (vs. a human) manager differently (multivariate $F$ (2, 216) = 0.40, $p$ = 0.67), either on procedural fairness ($t = -0.67$, $p = 0.50$, $BF_{01} = 18.18$), or on distributive fairness ($t = -0.88$, $p = 0.38$, $BF_{01} = 15.87$).

#### 4.2.4. Commitment

Future intentions to stay in or leave a managing system for a similar future task did not differ depending on the current algorithmic or human manager (see also Fig. 3B) – neither for serious ($t = -1.22$, $p = 0.22$, $BF_{01} = 14.71$) nor for unserious workers ($t = 1.03$, $p = 0.31$, $BF_{01} = 14.49$).

### 4.3. Discussion

Through pre-registered analyses, we did not find evidence for people's different reactions to algorithmic versus human management in our field experiment on Mturk. The awareness of being evaluated by an algorithmic or a human manager did not significantly influence people's performance or intrinsic motivation during the work process. Moreover, regardless of manager identity, fair or unfair treatment received equivalently positive or negative reactions, in terms of fairness perceptions and future intentions to stay in or leave the current managing system.

This experiment yielded novel findings that did not perfectly align with previous vignette or case studies. In contrast to what participants pessimistically predict in vignette studies (Acikgoz et al., 2020; Gonzalez et al., 2022; Lee, 2018; Raveendhran and Fast, 2021), workers in our field experiment did not show lower commitment under AI management, and did not feel AI managers to be less fair than human managers. In contrast to what employees optimistically report in case studies (Kusk and Bossen, 2022; Lix and Valentine, 2020; Sun et al., 2021), workers in our field experiment did not show higher performance or motivation under AI managers. In sum, what workers said and did in our field experiment stood somewhere in between the findings reported in vignette studies (where participants expect AI management to have negative effects on attitudes and behaviors) and the findings reported in case studies (where employees report positive attitudes and behaviors toward AI management).

## 5. General discussion

We argued that the two most prominent methods for studying the attitudes and behaviors of humans toward AI management, namely, vignette studies and case studies, often deliver inconsistent results due to their respective limitations. Vignette studies are likely to overestimate negative reactions to AI management due to affective forecasting errors, biased media coverage, and question substitution. Case studies may deliver too optimistic results given social desirability biases and lack of random assignment and control conditions. We suggested a third approach: field experiments on crowdsourced marketplaces, which can potentially overcome the limitations of vignette and case studies, and help consolidate previous findings from vignette and case studies. Behavioral scientists are often familiar with these platforms, using them as convenient subject pools – but they sometimes forget their primary nature as labor markets, which can be adapted to simulate work processes and organization under AI management.

We empirically demonstrated the unique advantages of such field experiments through a proof-of-concept experiment on Mturk, also in contrast to a representative-sample survey study. Since most people still have not experienced AI management in real life (Frank et al., 2019; Frey and Osborne, 2017; Huang and Rust, 2018), it was reasonable that people in the representative survey relied on their imaginations and indicated more negative reactions to AI management than they would actually experience. As expected, even though people expressed a strong aversion to AI managers (especially for managerial functions requiring subjective skills) due to autonomy and fairness concerns, these negative reactions did not emerge in our field experiment after people experienced AI management in a subjective task domain. This discrepancy also resonates with some other studies. For example, people anticipated that job interviews with AI managers would be less fair than job interviews with human managers (Acikgoz et al., 2020), but reported comparable feelings of fairness after simulated interviews with human and AI managers (Suen et al., 2019). In other words, a before-and-after contrast helped people recalibrate their negative expectations and adapt to systems with AI managers (Bucher et al., 2021; Curchod et al., 2020; Gonzalez et al., 2022; Haggadone et al., 2021; Makarius et al., 2020; Rahman, 2021).

Despite our attempts to summarize the respective limitations of vignette and case studies, the interpretation of each method and their findings should not be oversimplified. First, each method can have its unique advantages given applied contexts. For example, vignette studies can help anticipate worker reactions when organizations plan for a strategic transition from human to AI management, while case studies can better shed light on the narratives and social dynamics of a particular work group (e.g., Uber drivers; Langer and Landers, 2021). Second, within each methodological tradition, different studies can also have varied design decisions and nuanced findings. Some vignette studies strive to create immersive experience, and some case studies embed an experimental design. And they do not always yield conflicting findings; for example, vignette and case studies both reveal that people expect more consistency from AI-based than human decisions (Langer and Landers, 2021).

Different methodological approaches may not be the only reason why previous studies diverge on human reactions to AI management, and field experiments on crowdsourced marketplaces definitely come with their own limitations. For example, the results they deliver may not straightforwardly generalize to some contexts, such as workers in a dispatch center (Soper, 2021), drivers of ride-sharing services (Mohlmann and Zalmanson, 2018), or in-office workers (Jarrahi et al., 2021). More importantly, the novel approach may also induce novel ethical issues, which call for cautious interpretations of study findings and systematic regulations of the experiment procedure. For example, no difference between human and AI management should not be considered evidence that legitimizes undisclosed AI management without workers' awareness. And after experimental manipulations of (un)fair treatment by AI managers on crowdsourced marketplaces, workers should be debriefed about potential deception and receive justified payments (Wang et al., 2020).

With important caveats, we suggest that neither vignette nor case studies should be devalued, but their integration with our proposed third method – field experiments on crowdsourced marketplaces – can provide a better and more complete picture of human-centered AI management. Field experiments on crowdsourced marketplaces have the potential to help simulate the future of work, to make accurate predictions about workers' feelings and behaviors under AI management, and to indicate meaningful directions for technological innovations on AI-powered management.

## Ethics approval statement

The studies received ethics approval from the Max Planck Institute for Human Development (NO. C2021–3 and C2021-4).

## Funding statement

The authors acknowledge support from grant ANR-19-P13A-0004, grant ANR-17-EURE-0010, and the research foundation TSE-Partnership.

### Permission to reproduce material from other sources

There are no needs to reproduce materials from other sources.

### CRediT authorship contribution statement

**Mengchen Dong:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jean-François Bonnefon:** Conceptualization, Methodology, Validation, Visualization, Writing – review & editing. **Iyad Rahwan:** Conceptualization, Funding acquisition, Resources, Supervision, Validation, Writing – review & editing.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

The data can be found on Open Science Framework at https://tinyurl.com/AImanagement.

### Acknowledgements:

We thank Ard Barends for technical support, and Christopher Starke, Nils Köbis, and Zoe Purcell for valuable comments on the experiments and manuscript.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.technovation.2024.102953.

### References

Acikgoz, Y., Davison, K.H., Compagnone, M., Laske, M., 2020. Justice perceptions of artificial intelligence in selection. Int. J. Sel. Assess. 28 (4), 399–416. https://doi.org/10.1111/ijsa.12306.

Aguinis, H., Villamor, I., Ramani, R.S., 2021. MTurk research: review and Recommendations. J. Manag. 47 (4), 823–837. https://doi.org/10.1177/0149206320969787.

Arnold, H.J., Feldman, D.C., Purbhoo, M., 1985. The role of social-desirability response bias in Turnover research. Acad. Manag. J. 28 (4), 955–966. https://doi.org/10.5465/256249.

Bai, B., Dai, H., Zhang, D., Zhang, F., Hu, H., 2020. The impacts of algorithmic work assignment on fairness perceptions and productivity: evidence from field experiments. SSRN Electron. J. 1–32. https://doi.org/10.2139/ssrn.3550887.

Barends, A.J., de Vries, R.E., 2019. Noncompliant responding: comparing exclusion criteria in MTurk personality research to improve data quality. Pers. Indiv. Differ. 143 (February), 84–89. https://doi.org/10.1016/j.paid.2019.02.015.

Bigman, Y.E., Gray, K., 2018. People are averse to machines making moral decisions. Cognition 181. https://doi.org/10.1016/j.cognition.2018.08.003.

Bucher, E.L., Schou, P.K., Waldkirch, M., 2021. Pacifying the algorithm – anticipatory compliance in the face of algorithmic management in the gig economy. Organization 28 (1), 44–67. https://doi.org/10.1177/1350508420961531.

Buhrmester, M., Kwang, T., Gosling, S.D., 2011. Amazon's mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspect. Psychol. Sci. 6 (1), 3–5. https://doi.org/10.1177/1745691610393980.

Cao, G., Duan, Y., Edwards, J.S., Dwivedi, Y.K., 2021. Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making. Technovation 106, 102312. https://doi.org/10.1016/j.technovation.2021.102312.

Castelo, N., Bos, M.W., Lehmann, D.R., 2019. Task-dependent algorithm aversion. J. Market. Res. 56 (5), 809–825. https://doi.org/10.1177/0022243719851788.

Chandler, J., Shapiro, D., 2016. Conducting clinical research using crowdsourced convenience samples. Annu. Rev. Clin. Psychol. 12 (1), 53–81. https://doi.org/10.1146/annurev-clinpsy-021815-093623.

Colquitt, J.A., 2001. On the dimensionality of organizational justice: a construct validation of a measure. J. Appl. Psychol. 86 (3), 386–400. https://doi.org/10.1037/0021-9010.86.3.386.

Cormier, D., Newman, G., Nakane, M., Durocher, S., 2013. Would you do as a robot commands? An obedience study for human-robot interaction. In: International Conference on Human-Agent Interaction.

Crawford, K., 2021. Atlas of AI: the Real Worlds of Artificial Intelligence. YALE UNIVERSITY PRESS.

Curchod, C., Patriotta, G., Cohen, L., Neysen, N., 2020. Working for an algorithm: power asymmetries and agency in online work settings. Adm. Sci. Q. 65 (3), 644–676. https://doi.org/10.1177/0001839219867024.

Deci, E.L., Eghrari, H., Patrick, B.C., Leone, D.R., 1994. Facilitating internalization: the self-determination theory perspective. J. Pers. 62 (1), 119–142. https://doi.org/10.1111/j.1467-6494.1994.tb00797.x.

Dunning, D., Heath, C., Suls, J.M., 2004. Flawed self-assessment: implications for health, education, and the workplace. Psychol. Sci. Publ. Interest 5 (3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x.

Dzieza, J., 2020. How Hard Will the Robots Make Us Work? *the Verge*. https://www.theverge.com/2020/2/27/21155254/automation-robots-unemployment-jobs-vs-human-google-amazon.

Frank, M.R., Autor, D., Bessen, J.E., Brynjolfsson, E., Cebrian, M., Deming, D.J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., Rahwan, I., 2019. Toward understanding the impact of artificial intelligence on labor. Proc. Natl. Acad. Sci. U. S. A. 116 (14), 6531–6539. https://doi.org/10.1073/pnas.1900949116.

Frey, C.B., Osborne, M.A., 2017. The future of employment: how susceptible are jobs to computerisation? Technol. Forecast. Soc. Change 114, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019.

Geiger, A., 2016. Research in the Crowdsourcing Age, a Case Study. July 11. Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/2016/07/11/research-in-the-crowdsourcing-age-a-case-study/.

Gibson, W., 1999. Talk of the Nation, the Science in Science Fiction [National Public Radio]. http://www.npr.org/templates/story/story.php?storyId=1067220.

Glikson, E., Woolley, A.W., 2020. Human trust in artificial intelligence: review of empirical research. Acad. Manag. Ann. 14 (2), 627–660. https://doi.org/10.5465/annals.2018.0057.

Gombolay, M.C., Gutierrez, R.A., Clarke, S.G., Sturla, G.F., Shah, J.A., 2015. Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. Aut. Robots 39 (3), 293–312. https://doi.org/10.1007/s10514-015-9457-9.

Gonzalez, M.F., Liu, W., Shirase, L., Tomczak, D.L., Lobbe, C.E., Justenhoven, R., Martin, N.R., 2022. Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes. Comput. Hum. Behav. 130, 107179 https://doi.org/10.1016/j.chb.2022.107179.

Gray, H.M., Gray, K., Wegner, D.M., 2007. Dimensions of mind perception. Science 315 (5812), 619. https://doi.org/10.1126/science.1134475.

Haggadone, B.A., Banks, J., Koban, K., 2021. Of robots and robotkind: extending intergroup contact theory to social machines. Commun. Res. Rep. 38 (3), 161–171. https://doi.org/10.1080/08824096.2021.1909551.

Höddinghaus, M., Sondern, D., Hertel, G., 2021. The automation of leadership functions: would people trust decision algorithms? Comput. Hum. Behav. 116, 106635 https://doi.org/10.1016/j.chb.2020.106635.

Horton, J.J., Rand, D.G., Zeckhauser, R.J., 2011. The online laboratory: conducting experiments in a real labor market. Exp. Econ. 14 (3), 399–425. https://doi.org/10.1007/s10683-011-9273-9.

Huang, M.-H., Rust, R.T., 2018. Artificial intelligence in service. J. Serv. Res. 21 (2), 155–172. https://doi.org/10.1177/1094670517752459.

Jachimowicz, J.M., Duncan, S., Weber, E.U., Johnson, E.J., 2019. When and why defaults influence decisions: a meta-analysis of default effects. Behavioural Public Policy 3 (2), 159–186. https://doi.org/10.1017/bpp.2018.43.

Jarrahi, M.H., Newlands, G., Lee, M.K., Kinder, E., Sutherland, W., 2021. Algorithmic Management in a Work Context. Big Data & Society, Forthcomin. https://doi.org/10.1177/20539517211020332.

Kahneman, D., Frederick, S., 2002. Representativeness revisited: attribute substitution in intuitive judgment. In: Heuristics and Biases: the Psychology of Intuitive Judgment. Cambridge University Press, pp. 49–81. https://doi.org/10.1017/CBO9780511808098.004.

Kawaguchi, K., 2020. When Will Workers Follow an Algorithm? A Field Experiment with a Retail Business. Management Science. https://doi.org/10.1287/mnsc.2020.3599.

Kellogg, K.C., Valentine, M.A., Christin, A., 2020. Algorithms at work: the new contested terrain of control. Acad. Manag. Ann. 14 (1), 366–410. https://doi.org/10.5465/annals.2018.0174.

Köbis, N., Mossink, L.D., 2021. Artificial intelligence versus Maya Angelou: experimental evidence that people cannot differentiate AI-generated from human-written poetry. Comput. Hum. Behav. 114, 106553 https://doi.org/10.1016/j.chb.2020.106553.

Kraus, M.W., Horberg, E.J., Goetz, J.L., Keltner, D., 2011. Social class rank, threat vigilance, and hostile reactivity. Pers. Soc. Psychol. Bull. 37 (10), 1376–1388. https://doi.org/10.1177/0146167211410987.

Kusk, K., Bossen, C., 2022. Working with Wolt: an ethnographic study of lenient algorithmic management on a food delivery platform. In: Proc. ACM Hum.-Comput. Interact., vol. 6. Association for Computing Machinery. Issue GROUP, p. Article 4).

Langer, M., Landers, R.N., 2021. The future of artificial intelligence at work: a review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. Comput. Hum. Behav. 123, 106878 https://doi.org/10.1016/j.chb.2021.106878.

Lee, M.K., 2018. Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management. Big Data Soc. 5 (1), 1–16. https://doi.org/10.1177/2053951718756684.

Lee, K., Ashton, M.C., 2018. Psychometric properties of the HEXACO-100. Assessment 25 (5), 543–556. https://doi.org/10.1177/1073191116659134.

Lee, M.K., Nigam, I., Zhang, A., Afriyie, J., Qin, Z., Gao, S., 2021. Participatory algorithmic management: elicitation methods for worker well-being models. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. Association for Computing Machinery, pp. 715–726. https://doi.org/10.1145/3461702.3462628.

Lix, K., Valentine, M.A., 2020. When a Bot Scores Your Karma: Algorithmic Rankings as Uncertainty Reducers in Platform Gig Work Poses New Management Challenges.

Logg, J.M., Minson, J.A., Moore, D.A., 2019. Algorithm appreciation: people prefer algorithmic to human judgment. Organ. Behav. Hum. Decis. Process. 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005. April 2018.

Longoni, C., Bonezzi, A., Morewedge, C.K., 2019. Resistance to medical artificial intelligence. J. Consum. Res. 46 (4), 629–650. https://doi.org/10.1093/jcr/ucz013.

Lount, R.B., Pettit, N.C., 2012. The social context of trust: the role of status. Organ. Behav. Hum. Decis. Process. 117 (1), 15–23. https://doi.org/10.1016/j.obhdp.2011.07.005.

Lucas, G.M., Gratch, J., King, A., Morency, L.-P., 2014. It's only a computer: virtual humans increase willingness to disclose. Comput. Hum. Behav. 37, 94–100. https://doi.org/10.1016/j.chb.2014.04.043.

Makarius, E.E., Mukherjee, D., Fox, J.D., Fox, A.K., 2020. Rising with the machines: a sociotechnical framework for bringing artificial intelligence into the organization. J. Bus. Res. 120, 262–273. https://doi.org/10.1016/j.jbusres.2020.07.045.

MIT Work of the Future, 2019. The Work of the Future: Shaping Technology and Institutions. https://workofthefuture.mit.edu/wp-content/uploads/2020/08/WorkoftheFuture_Report_Shaping_Technology_and_Institutions.pdf.

Mohlmann, M., Zalmanson, L., 2018. Hands on the wheel: navigating algorithmic management and uber drivers' autonomy. In: ICIS 2017: Transforming Society with Digital Innovation. December.

Morewedge, C.K., 2022. Preference for human, not algorithm aversion. Trends Cognit. Sci. https://doi.org/10.1016/j.tics.2022.07.007. S1364661322001644.

Moss, A., 2020. Demographics of People on Amazon Mechanical Turk. CloudResearch. https://www.cloudresearch.com/resources/blog/who-uses-amazon-mturk-2020-demographics/.

Newman, D.T., Fast, N.J., Harmon, D.J., 2020. When eliminating bias isn't fair: algorithmic reductionism and procedural justice in human resource decisions. Organ. Behav. Hum. Decis. Process. 160, 149–167. https://doi.org/10.1016/j.obhdp.2020.03.008. April 2018.

Paolacci, G., Chandler, J., Ipeirotis, P.G., 2010. Running experiments on Amazon mechanical turk. Judgment and Decision Making 5 (5), 411–419.

Parker, S.K., Grote, G., 2022. Automation, algorithms, and beyond: why work design matters more than ever in a digital world. Appl. Psychol. 71 (4), 1171–1204. https://doi.org/10.1111/apps.12241.

Rahman, H.A., 2021. The invisible cage: workers' reactivity to opaque algorithmic evaluations. Adm. Sci. Q. 66 (4), 945–988. https://doi.org/10.1177/00018392211010118.

Ranganathan, A., Benson, A., 2020. A numbers game: quantification of work, auto-gamification, and worker productivity. Am. Socio. Rev. 85 (4), 573–609. https://doi.org/10.1177/0003122420936665.

Raveendhran, R., Fast, N.J., 2021. Humans judge, algorithms nudge: the psychology of behavior tracking acceptance. Organ. Behav. Hum. Decis. Process. 164, 11–26. https://doi.org/10.1016/j.obhdp.2021.01.001.

Rucker, D.D., Galinsky, A.D., Magee, J.C., 2018. The agentic–communal model of advantage and disadvantage: how inequality produces similarities in the psychology of power, social class, gender, and race. In: Advances in Experimental Social Psychology, vol. 58. Elsevier, pp. 71–125. https://doi.org/10.1016/bs.aesp.2018.04.001.

Soper, S., 2021. Fired by Bot at Amazon: 'It's You against the Machine.' *Bloomberg*. https://www.bloomberg.com/news/features/2021-06-28/fired-by-bot-amazon-turns-to-machine-managers-and-workers-are-losing-out.

Suen, H.-Y., Chen, M.Y.-C., Lu, S.-H., 2019. Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? Comput. Hum. Behav. 98, 93–101. https://doi.org/10.1016/j.chb.2019.04.012.

Sun, J., Zhang, D.J., Hu, H., Van Mieghem, J.A., 2021. Predicting human discretion to adjust algorithmic prescription: a large-scale field experiment in warehouse operations. Manag. Sci. 3990. https://doi.org/10.1287/mnsc.2021.3990 mnsc.2021.

Tong, S., Jia, N., Luo, X., Fang, Z., 2021. The Janus face of artificial intelligence feedback: deployment versus disclosure effects on employee performance. Strat. Manag. J. 42 (9), 1600–1631. https://doi.org/10.1002/smj.3322.

von der Pütten, A.M., Krämer, N.C., Gratch, J., Kang, S.-H., 2010. "It doesn't matter what you are!" Explaining social effects of agents and avatars. Comput. Hum. Behav. 26 (6), 1641–1650. https://doi.org/10.1016/j.chb.2010.06.012.

Wang, R., Harper, F.M., Zhu, H., 2020. Factors influencing perceived fairness in algorithmic decision-making: algorithm outcomes, development procedures, and individual differences. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–14. https://doi.org/10.1145/3313831.3376813.

Wesche, J.S., Sonderegger, A., 2019. When computers take the lead: the automation of leadership. Comput. Hum. Behav. 101, 197–209. https://doi.org/10.1016/j.chb.2019.07.027.

White, J.C., Ravid, D.M., Siderits, I.O., Behrend, T.S., 2022. An urgent call for IO psychologists to produce timelier technology research. Ind. Org. Psychol. 15 (3), 441–459. https://doi.org/10.1017/iop.2022.26.

Wilson, T.D., Gilbert, D.T., 2005. Affective forecasting: knowing what to want. Curr. Dir. Psychol. Sci. 14 (3), 131–134. JSTOR.

Wood, A., 2021. Algorithmic Management: Consequences for Work Organisation and Working Conditions (No. JRC124874; Seville: European Commission). https://ec.europa.eu/jrc/sites/default/files/jrc124874.pdf.

Wood, S., Michaelides, G., Inceoglu, I., Niven, K., Kelleher, A., Hurren, E., Daniels, K., 2022. Satisfaction with One's Job and Working at Home in the COVID-19 Pandemic: A Two-wave Study. Appl. Psychol.,, 12440. https://doi.org/10.1111/apps.12440 apps.