frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 25 May 2022
doi: 10.3389/fpsyg.2022.779028

Blame the Machine? Insights From an Experiment on Algorithm Aversion and Blame Avoidance in Computer-Aided Human Resource Management

Christian Maasland[1†] *and Kristina S. Weißmüller*[2*†]

[1]Independent Researcher, Hamburg, Germany, [2]KPM Center for Public Management, University of Bern, Bern, Switzerland

OPEN ACCESS

Edited by:
James Gaskin,
Brigham Young University,
United States

Reviewed by:
Ludivine Martin,
Luxembourg Institute of Socio-Economic Research,
Luxembourg
Kristian Rotaru,
Monash University, Australia
Paweł Niszczota,
Poznań University of Economics and Business, Poland

***Correspondence:**
Kristina S. Weißmüller
kristina.weissmueller@kpm.unibe.ch

[†]These authors share first authorship

Specialty section:
This article was submitted to Organizational Psychology, a section of the journal Frontiers in Psychology

Received: 17 September 2021
Accepted: 11 April 2022
Published: 25 May 2022

Citation:
Maasland C and Weißmüller KS (2022) Blame the Machine? Insights From an Experiment on Algorithm Aversion and Blame Avoidance in Computer-Aided Human Resource Management. Front. Psychol. 13:779028. doi: 10.3389/fpsyg.2022.779028

Algorithms have become increasingly relevant in supporting human resource (HR) management, but their application may entail psychological biases and unintended side effects on employee behavior. This study examines the effect of the type of HR decision (i.e., promoting or dismissing staff) on the likelihood of delegating these HR decisions to an algorithm-based decision support system. Based on prior research on algorithm aversion and blame avoidance, we conducted a quantitative online experiment using a 2×2 randomly controlled design with a sample of $N = 288$ highly educated young professionals and graduate students in Germany. This study partly replicates and substantially extends the methods and theoretical insights from a 2015 study by Dietvorst and colleagues. While we find that respondents exhibit a tendency of delegating presumably unpleasant HR tasks (i.e., dismissals) to the algorithm—rather than delegating promotions—this effect is highly conditional upon the opportunity to pretest the algorithm, as well as individuals' level of trust in machine-based and human forecast. Respondents' aversion to algorithms dominates blame avoidance by delegation. This study is the first to provide empirical evidence that the type of HR decision affects algorithm aversion only to a limited extent. Instead, it reveals the counterintuitive effect of algorithm pretesting and the relevance of confidence in forecast models in the context of algorithm-aided HRM, providing theoretical and practical insights.

Keywords: algorithm aversion, blame avoidance, human resource management, algorithm-based decision support systems, behavioral experimental research

INTRODUCTION

With the rise of people analytics, human resource (HR) management today relies heavily on algorithm-based decision support systems to assist HR managers in the assessment of the value of individual employees as part of their organizations' human capital asset (Reindl, 2016; Leicht-Deobald et al., 2019). People analytics is the creation of unique workforce insights by integrating originally disparate data sources from both inside and outside the organization to create HR insights with strategic value that lead to a competitive advantage.

While metric-driven analytical approaches to HR management date back to the beginning of the 20th century (Huselid, 1995; Leicht-Deobald et al., 2019), quantifying individuals' human capital as a specific asset to the firm has become much more sophisticated because of recent technological advances in algorithm-based machine learning and big data (Reindl, 2016). In a 2017 survey of 10,400 firms across 140 countries, Deloitte identified people analytics as one of the main global trends in human capital, with 71% of companies surveyed rating algorithm-based people analytics as a high priority in their organizations (Fineman et al., 2017). Today, a diverse landscape of software provides algorithm-based people analytic solutions for HR across all industries. For instance, Xerox Services have been using algorithm-based people analytics since as early as 2010 (Peck, 2013). Other key players in personnel information systems such as IBM, SAP, and Oracle use integrated application tools to accumulate HR data from existing databases (Angrave et al., 2016). Typical case studies of firms using algorithms strategically to enhance the efficiency of their talent management are, among many others, the tech giants Google (*People Analytics*; Shrivastava et al., 2018) and Microsoft (*MyAnalytics*; Giermindl et al., 2021), the bank ING (Peeters et al., 2020), the cybersecurity firm Juniper Networks (Boudreau and Rice, 2015), the retailer Wal-Mart (Haube, 2015), and online retailer Zalando using the software *Zonar* (Staab and Geschke, 2020).

Since the aforementioned workforce insights—generated by scoring employees based on computer-aided procedures—may inform HR practices not only descriptively but may also be used to predict future performance, applying people analytics raises substantial questions regarding legal and ethical matters (Leicht-Deobald et al., 2019), particularly regarding the quality of algorithmic predictions and their effect on HR managers' choice architectures (Reindl, 2016). The idea of enhancing the quality of strategic HR management by using IT-based support systems that facilitate personnel decisions is deeply rooted in the paradigm of evidence-based management (Sharma and Sharma, 2017; Leicht-Deobald et al., 2019). In fact, the discourse on mechanical vs. clinical decision making shows that algorithms often outperform human forecast accuracy (Meehl, 1954; Sawyer, 1966; Kuncel et al., 2013). Yet, decision makers feel ambiguous about using algorithm-based support systems even though they know that doing so would optimize their choices and lead to objectively better outcomes (Grove and Meehl, 1996; Sanders and Manrodt, 2003; Fildes and Goodwin, 2007; Highhouse, 2008). Due to its relevance and rising popularity with decision makers in HRM (see, e.g., Sharma and Sharma, 2017), behavioral research on the effects of the availability of algorithm-based decision support systems on HR-related decision is highly relevant and has become a rapidly growing field of research lately (Burton et al., 2020).

As one of the first experimental studies in the field of algorithm aversion, Dietvorst et al. (2015) found that decision makers were reluctant to let their choice be guided by an algorithm-aided forecast model to determine which MBA students should be granted a scholarship although the quality and accuracy of the model forecast was clearly superior to human forecast ability. This phenomenon of *algorithm aversion* is intriguing because it contradicts the classic assumptions of utility maximizing behavior. While Dietvorst et al. (2015, 2018) successfully replicated their initial experiment, other studies further explored the effects of algorithm familiarity and trust (Berger et al., 2021; Filiz et al., 2021), algorithms' characteristics such as their ability to learn (Berger et al., 2021), and their perceived fairness (Newman et al., 2020) as determinants of algorithm aversion, the underlying psychological mechanisms of algorithm aversion are still not explored sufficiently enough yet. Particularly, the effect of different choice types and valences on algorithm aversion is not well understood yet (but see initial findings by Newman et al., 2020, and Renier et al., 2021). Both studies by Dietvorst et al. (2015, 2018) are framed in a positive scenario of making performance evaluations with presumably rather positive valence (e.g., selecting students for a scholarship), but behavioral research on risk preferences (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991) and blame avoidance in contextual frames (Vis and van Kersbergen, 2007; Bartling and Fischbacher, 2012) suggest that boundedly rational decision makers would react very differently when faced with the task of making decisions generally presumed to be unpleasant such as taking away the scholarship or dealing punishment. Little is known about whether people may prefer to delegate unpleasant choices to an algorithm to avoid blame and negative sentiment by shifting their personal responsibility to the machine (Langer and Landers, 2021). This is particularly relevant for HR management since HR managers regularly face tough workforce-related decisions—e.g., whom to lay off and whom to grant a promotion—while often being personally involved with their workforce. It is an obvious assumption that HR managers may be more likely to delegate making unpleasant HR decisions—e.g., laying off employees—to an algorithm compared to making more pleasant decisions (e.g., promoting employees) because the latter may entail less emotional burden and may offer the hedonic utility of making pleasant HR choices (Hamman et al., 2010). Does the type of HR choice task affect algorithm aversion?

The current study reports quantitative evidence from an original between-subject experimental study on the conditionality of algorithm aversion in positive and negative valence settings. Specifically, we report evidence from a 2×2 randomized controlled online vignette experiment conducted with a sample of $N = 288$ highly educated German residents. Set in the context of strategic HR management, we replicate prior experimental research by Dietvorst et al. (2015, 2018) and enhance their original experimental design by adding a contextual positive vis-à-vis negative affective vignette-based treatment (*personnel promotion* vs. *personnel dismissal*) to test whether decision makers are more likely to use algorithm-based decision support systems for making presumably unpleasant decisions—i.e., laying off staff—compared to presumably more pleasant decisions—i.e., promoting staff. We also control for the perceived algorithmic forecast precision, the role of pretesting it, and individuals' confidence in human (CIH) and machine (CIM) forecast.

Following explicit calls by Dietvorst et al. (2015, 2018), Prahl and van Swol (2017), Tambe et al. (2019), and Newman et al. (2020), our research design comes with a few crucial

methodological advantages. Its experimental setup with randomized controlled trials allows us to identify the latent causal mechanisms that relate algorithm aversion to blame avoidance based on the valence frame of a choice situation (i.e., promoting or dismissing staff). Our findings are directly relevant for HR management in practice because they add a new perspective to the scientific discourse on bounded rationality in the age of computer-aided choice architectures. This allows us to offer advice to HR managers in the form of caveats when using algorithms to enhance quality and precision in HR management.

In the next section, we review the literature on the motivational foundations of algorithm aversion and blame avoidance and derive two hypotheses on individuals' likelihood to delegate critical HR decisions to algorithms in relation to individuals' CIH and machine judgment. Then, we present the experimental design and procedure and report the results of the hypothesis testing with experimental data from $N = 288$ highly educated German respondents. We conclude with a discussion of the implications of our findings for theory and practice and suggest avenues for future research.

THEORY

Origins of Algorithm Aversion

Algorithms are generally defined as a set of mathematical instructions that—without explicit human intervention—help calculate a solution to a given problem (Lee, 2018). Algorithms are based on elaborate statistical techniques that result in sophisticated forecasting models. In HR management, algorithm-based decision support systems profit from recent technical developments in machine learning and artificial intelligence that allow for high-level automatization in decision making "to supplement and inform (and perhaps supersede) human judgment or intuition" (Dietvorst et al., 2015; Prahl and van Swol, 2017; Lee, 2018).

Yet, empirical research across the whole spectrum of management science shows that decision makers are reluctant to use algorithms to maximize forecast precision and that people tend to discount machine-based forecast—compared to human-made forecast—even if explicitly informed about its superiority (Fildes and Hastings, 1994; Mentzer and Kahn, 1995; Dzindolet et al., 2002; Sanders and Manrodt, 2003; Fildes and Goodwin, 2007). For instance, Önkal et al. (2009) conducted a framing experiment with 130 graduate business administration students who were asked to predict stock prices. The experiment revealed that study participants discounted the perceived accuracy of a prediction presented as algorithm-based advice much more steeply compared to the case in which the very same prediction was presented as human-made. Medical, psychological, and financial studies also show that people generally prefer human judgment over machine-aided models of prediction and find human forecasts more trustworthy (Lee, 2018; Filiz et al., 2021). Research on computer-aided decision processes by Dzindolet et al. (2002) and Renier et al. (2021) shows that decision makers tend to perceive observed error rates of algorithm-based models as disproportionately more negative compared to human estimate-based models presumably because machine-made errors are incongruent to the widely held idea of perfection in machine-made forecast (Stangor and McMillan, 1992; Madhavan and Wiegmann, 2007; Boyd and Crawford, 2012; Constantiou and Kallinikos, 2015; Berger et al., 2021) and because negative information cues are psychologically more salient than positive information cues (Rozin and Royzman, 2001).

Prior studies on algorithm aversion hypothesized that there are a number of reasons for this preference for human-made forecast even in spite of explicit superiority of choices made by an algorithm: for instance, the notion that using algorithm-based choice modeling may be perceived as a loss of process ownership (Önkal and Gönül, 2005; Petropoulos et al., 2016), an abstract sense of unfamiliarity and hence distrust with the machine (Prahl and van Swol, 2017), or the notion that algorithms were unable to integrate qualitative factors (Grove and Meehl, 1996; Vrieze and Grove, 2009; Newman et al., 2020). Others argue that decision makers perceive algorithms as unable to account for unique and individual circumstances (Highhouse, 2008; Longoni et al., 2019), unable to deliver satisfying results in domains of high uncertainty (Dietvorst and Bharti, 2020), or mention machines' lack of intuition and fairness (Newman et al., 2020), a quality typically associated with human forecasting (Lee, 2018; Burton et al., 2020). However, Dietvorst et al. (2015, 2018) were the first to conduct a series of experimental studies to identify the underlying behavioral mechanisms of algorithm aversion, especially concerning the correlation between the perceived fallibility of the algorithm-based choice models and the likelihood of delegating decisions to this algorithm. They found that decision makers were significantly less likely to delegate to the machine after seeing it perform (and inevitably err) even though the algorithm still dramatically outperformed their human judgment (Dietvorst et al., 2015, 2018). This effect was independent of the incentive structure, and their findings were replicated in follow-up studies by Prahl and van Swol (2017) and Dietvorst and Bharti (2020).

Recent studies have analyzed factors that may help reduce algorithm aversion. For instance, granting users limited control over the algorithm's forecast reduced algorithm aversion (Dietvorst et al., 2018), and people tended to pardon an algorithm's error if it was small and the decision domain was relatively predictable (Dietvorst and Bharti, 2020). Lee (2018) analyzed the acceptance rate of algorithm-based choices for tasks that require mechanical as opposed to human skills, and Castelo et al. (2019) found that increasing task objectivity and the human semblance of an algorithm's pattern of decision making lead to higher trust in the algorithm. Yet to date, the underlying behavioral mechanisms of algorithm aversion regarding distinct types of decisions are still unexplored.

Responsibility Shifting and Blame Avoidance

Prior studies exploring how people cope with making challenging or nonsocially acceptable decisions showed that by delegating a decision, individuals indeed shift both the mental burden

and the factual or perceived responsibility for making the decision. Bartling and Fischbacher (2012) and Oexl and Grossman (2013) examined this psychological shifting process and showed that individuals affected by the outcome of a decision will in fact not blame the person responsible for making the decision but the person executing and delivering it. They stress that blame avoidance by shifting responsibility is a major motive for delegating unpopular decisions (see also Hill, 2015). Another example is the study by Erat (2013) that revealed that people will deliberately delegate the act of lying to their subordinates to avoid the responsibility for lying because such behavior is associated with negative sentiment, arousing psychological burdens in the form of hedonic disutility and the risk of blame. This finding corresponds with prior studies showing that individuals seek to avoid situations in which they may harm others, because decision delegation reduces decision makers' mental costs of feeling responsible (Steffel et al., 2016). These psychological and moral factors relating to accountability are important predictors of algorithm aversion (Giermindl et al., 2021). For instance, experimental research by Newman et al. (2020) shows that people perceive algorithm-made HR decisions as less fair, a finding that was stable irrespective of whether employees were selected for promotion or layoff.

Given that the discourse on algorithm aversion suggests that delegating to an algorithm reduces perceived process ownership (Önkal and Gönül, 2005; Petropoulos et al., 2016), and given that the discourse on blame avoidance suggests reduced ownership is a behavioral strategy to cope with mental burdens in challenging choice situations, we hypothesize that the likelihood of delegating a HR decision to an algorithm is task dependent in the sense that:

> *Hypothesis 1 (H1)*: Decision makers are more likely to delegate the decision of dismissing employees to an algorithm compared with promoting employees.

Yet, prior research on algorithm aversion (Dietvorst et al., 2015, 2018; Dietvorst and Bharti, 2020) indicates that individuals will be reluctant to delegate decision making to an algorithm if they feel its forecasting error is too high, rendering it unreliable. Making a HR decisions that may potentially change an employee's life is a challenging situation, particularly for diligent HR managers. In the era of people analytics, HR decision makers are faced with peculiar moral conflict (Tambe et al., 2019): weighing the potential benefits and costs of delegating to an algorithm-based decision support system reduces individual mental costs by reducing perceive accountability and subjective expected blame but knowingly using a flawed algorithm may pose a violation of decision makers' moral concept of self as an ethical and virtuous (i.e., "*blameless*") person—particularly since many people assume that algorithm-based HR decision are less fair (Newman et al., 2020). The preservation of one's moral concept of self is a psychological motive with high priority that strongly affects choice behavior (Bem, 1972; Hsee, 1996). Pretesting any genuine algorithm will reveal the realistic limits of its predictive quality and, hence, decision makers' capacity to justify using it because "imperfect" algorithms violate the widely held expectation of software infallibility (Boyd and Crawford, 2012; Dietvorst and Bharti, 2020). This is why, among others, Dietvorst and Bharti (2020) and Renier et al. (2021), suggest that pretesting an algorithm reduces the likelihood of using it, suggesting that:

> *Hypothesis 2a (H2a)*: Pretesting an algorithm reduces the likelihood of delegating HR decisions to the algorithm.

However, the perceived costs associated with delegating to a pretested algorithm may yet be conditional upon the type and valence of the choice to be delegated (Langer and Landers, 2021). While Newman et al. (2020) found that HR decisions made by an algorithm were perceived as less fair in both promotion and layoff decisions, the effect size of perceived human–machine difference in decision quality was lower for negative valence (i.e., layoffs) vis-à-vis positive valence (i.e., promotion) choice tasks. This points toward a potential asymmetry of the pretesting-related negativity bias toward the algorithm. However, recent empirical studies show that the degree of decision makers' reaction to pretesting an algorithm may be conditional upon the quality of this pretesting experience. The perceived relative precision of the machine vis-à-vis human forecast precision may influence decision makers' response (Filiz et al., 2021), as well as the specific task characteristic, particularly if stakes are high with regards to both tangible and moral costs (e.g., in the form of mental burdens; Lee, 2018; Burton et al., 2020). While research into this nexus is yet inconclusive (Langer and Landers, 2021), Renier et al. (2021) reveal that algorithmic fallibility will trigger harsher and comparatively more negative psychological reactions compared to experiencing equivalent human error and that a high-stakes HR choice context may be particularly salient in determining the acceptability of using an "imperfect" algorithm. Taken together, these initial findings on topical asymmetries suggest an alternative hypothesis.

> *Hypothesis 2b (H2b)*: Pretesting the algorithm reduces the relative likelihood of delegating the decision of dismissing employees to an algorithm compared with promoting employees.

MATERIALS AND METHODS

Experimental Design and Sample

The current study investigates whether people are more likely to delegate HR decisions to a computer algorithm if their decision is related to laying off employees compared to promoting employees. To test our hypotheses, we conducted a quantitative study using an interactive and dynamic online experiment in a randomized controlled 2×2 vignette design following best-practice advice by Atzmüller and Steiner (2010) and Aguinis and Bradley (2014) to warrant high levels of internal and external validity and to minimize social desirability-related

response bias (Fisher and Katz, 2000). Based on actual HR decision support systems, we designed two equivalent yet topically opposite vignette scenarios—one dealing with making decisions on *promoting employees* (P) and one dealing with *dismissing employees* (D)—both of which offered respondents the opportunity to delegate the decision to a computer algorithm specifically designed to support these tasks. The two experimental conditions were corroborated with two alternative treatments to replicate algorithm aversion experiment of Dietvorst et al. (2015). The first treatment encompasses *pretesting* the algorithm before making the delegation choice (*pretest*; pt). The second treatment offered *no* option for *pretesting* the algorithm (nt). Consequently, the experiment defines four treatment groups ($P_{\_nt}$, $D_{\_nt}$, $P_{\_pt}$, and $D_{\_pt}$) based on two independent stimuli (choice task vignette condition *P* or *D* and treatment *nt* or *pt*) and a single dependent variable *delegation choice*, i.e., the decision of delegating the HR decision to the algorithm.

The online experiment was conducted between March and May 2018 and took between 20 and 30 min to complete. A convenience sample was raised by distributing the link to the experiment *via* several universities' e-mail lists addressed to young professionals and graduate students and through online career networks, reaching a total of 574 individuals of which $N = 288$ (50.2%) fully completed the experiment. Respondents were incentivized by the prospect of winning one of several gift vouchers (€25, €50, or €75) for a popular online retailer. To warrant rigor, only complete responses were included in the final dataset of this study. The final sample comprises 156 (54.2%) women, 115 (39.9%) men, and 17 (5.9%) individuals who did not disclose. Respondents are on average $M = 28.03$ ($SD = 6.1$) years old, predominantly German citizens (91.0%), and highly educated with 169 (58.7%) having completed a tertiary degree education.

Experimental Procedure

Figure 1 provides an overview of the experimental procedure. Since the current study partially replicates the experiments conducted by Dietvorst et al. (2015, 2018), our vignette design and treatment wording were designed to resemble the former studies' procedures as closely as possible. The original procedures were enhanced by adding the two different HR-related task stimuli—i.e., *promotion* or *dismissal*—to the vignette treatment, resulting in the 2×2 design.

First, respondents were introduced to the experimental setting placed in the context of employee performance evaluation. Participants were randomly sorted into one of the four treatment conditions and introduced to their role and respective tasks in the vignette scenario of this study. Respondents were asked to imagine being the *Chief Information Officer* (CIO) of a big company that employs hundreds of software engineers and to evaluate how successful current software engineers might be as software consultants in the future in comparison to other employees of the company's workforce. In the promotion vignette, respondents were informed that their task was to predict the future performance of ten employees based on eight carefully selected criteria typical for software engineers. They were told that these employees would later be ranked against each other and that all employees who scored above a certain but unknown threshold would be promoted.[1] Similarly, in the dismissal vignette, respondents presented with the identical criteria and information were informed that their predictions regarding employees' future performance was to be used to determine that employees who scored below a certain unknown threshold level would be dismissed. These conditions provide a balanced equivalent treatment framed in two different choice scenarios.

Next, all participants were informed that they were free to use an algorithm-based decision support system, which would produce forecasts based on the very same data available to the respondents themselves. Participants were given further information to confirm that it was a sophisticated algorithm created by diligent expert analysts.

Then, study participants were, again, randomly sorted into the *pretest* (pt) and the *no-test* (nt) condition branch of the experiment. Respondents in the pretest condition were asked to make 10 practice rounds, in which they could see the algorithm perform before they would make the 10 employee predictions, which would also determine their likelihood of winning the incentive lottery. In each of the 10 trial rounds, respondents were informed about employees' "real" performance score (range: 1–100) and the performance score predicted by the algorithm, respectively. Respondents sorted into the no-test condition directly moved on to their task, i.e., they had no chance to see the algorithm perform *a priori*.[2] In this context, it is important to note that the algorithm's absolute average prediction errors (AAE) across all trials were designed following procedure of Dietvorst et al. (2015) to ensure that human and algorithmic forecast accuracy were virtually identical across all conditions, vignettes, and rounds. This component is an important design aspect to warrant that there is no quality-related, functional reason to disregard algorithm support. The main dependent variable of this experiment is whether respondents choose to make the forecast themselves (*delegation choice* = 0) or delegate their task to the algorithm (*delegation choice* = 1). Respondents were randomly presented with 10 vignettes drawn from the set of 20 employee vignettes and asked to predict each employee's future performance on a scale from 1 to 100. Prior to these 10 rounds, respondents were asked whether they would like to tie their predictions to those made by the algorithm. During these 10 rounds, participants saw no information about employees' "true" performance scores. After completing the 10 rounds, respondents were asked to indicate their confidence in human forecast and their confidence in the algorithm forecast, respectively, on a five-point Likert-type scale ranging from 1 = "*no confidence*" to 5 = "*full confidence*." We use this *post-hoc* measure for explorative analyses.

[1]These criteria were carefully designed to optimize the external validity of the experiment by maximizing respondents' immersion into a highly realistic choice scenario. A more extensive description of the experiment and stimuli is presented in **Appendix A** and **B**.

[2]To inhibit halo and primacy effects (Shteingart et al., 2013), the employee-based stimuli were presented in random order between subjects and were drawn from a carefully designed sample of 20 software engineers to introduce sufficient amounts of variance (see also **Appendix C**).

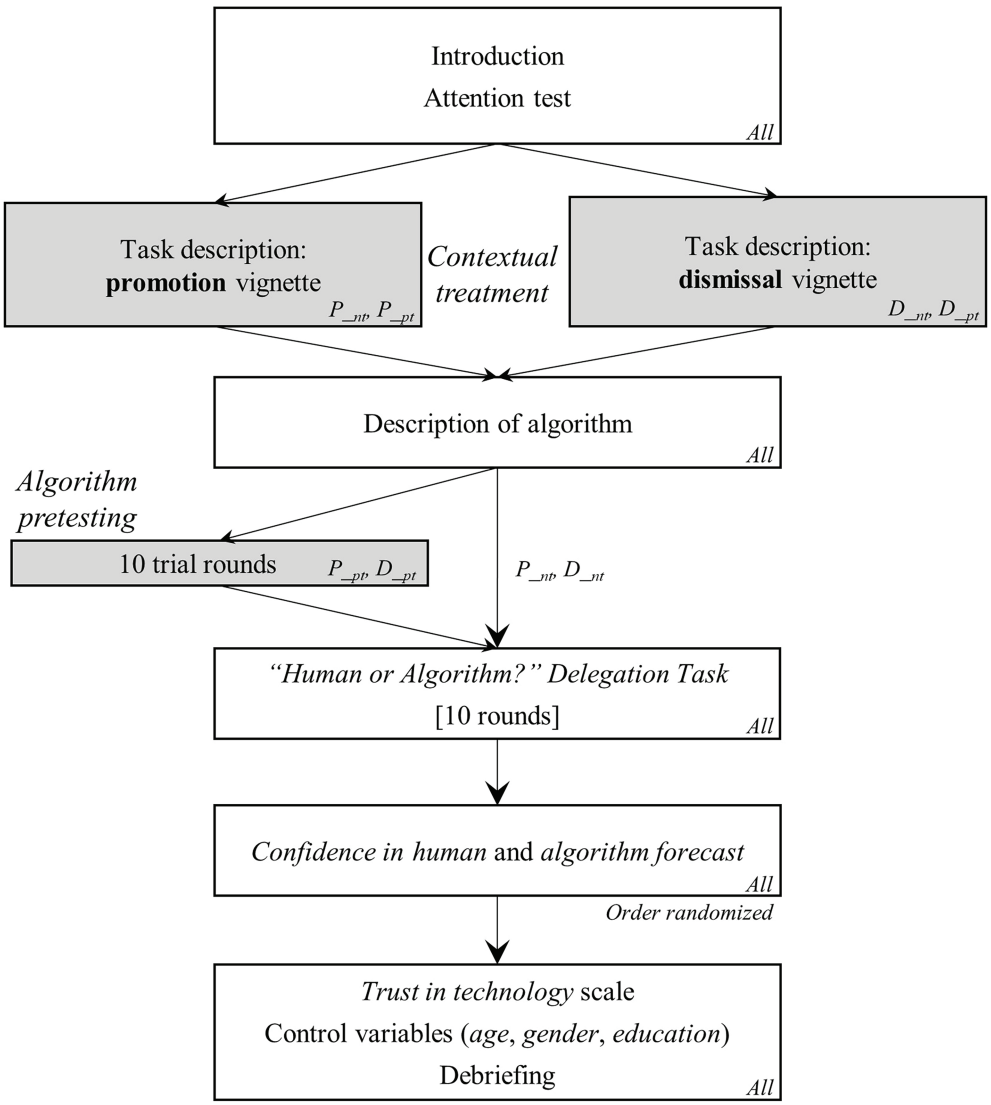

FIGURE 1 | Experimental procedure.

As a control variable, respondents' *trust in technology* was measured after to the experimental tasks with seven-item Likert-type measure of McKnight et al. (2011) on trust in information technology. The original English scale items were translated into German with due diligence. Furthermore, respondents were asked to indicate their *age*, *gender*, *level of education*, and *country of residence*, in each case allowing for nonresponse, before being debriefed.

RESULTS

Descriptive Analysis

The four vignette treatments were randomly distributed among the study participants. As not all respondents finished the experiment, the treatment distribution varies; the overall distributions of treatments are $n_{P\_nt} = 77$ (26.7%), $n_{D\_nt} = 81$ (28.1%), $n_{P\_pt} = 59$ (20.5%), and $n_{D\_pt} = 71$ (24.7%). Of the final sample, 136 (47.2%) received the positive *promotion* frame (P) and 152 (52.8%) of the final sample received the negative *dismissal* frame (D), 158 (54.9%) people received the no-test condition (nt) and 130 (45.1%) the pretest condition (pt).

Table 1 displays the correlation matrix of all variables. For the current sample, the bidimensional *trust in technology* measure of McKnight et al. (2011) resulted in a satisfying level of construct reliability [faith in technology (general): $\alpha_{fit} = 0.61$; trusting stance: $\alpha_{ts} = 0.69$]. Of all $N = 288$ respondents, $n = 66$ (23.0%) chose to use the algorithm. Logistic regression modeling with [$\chi^2(8) = 55.29$, $p < 0.000$; pseudo-$R^2 = 0.200$] and without control variables [$\chi^2(4) = 56.78$, $p < 0.000$; pseudo-$R^2 = 0.195$] reveals that neither respondents' trust in technology, age, gender, nor education explained any substantial amount of variance in choice. However, individuals' confidence in human and

TABLE 1 | Correlation matrix.

	Variable	1	2	3	4	5	6	7	8
1	Delegation choice [0, 1] = [human decision; algorithm's decision]	–							
2	Algorithm pretested? [0, 1] = [no; yes]	-0.18_a**	–						
3	HR choice task [0, 1] = [promotion; dismissal]	0.05_a	0.03_a	–					
	Confidence in…								
4	…Algorithm forecast	0.36_a***	-0.15_a*	-0.08_a	–				
5	…Human forecast	-0.25_a***	-0.02_a	-0.04_a	-0.12_c*	–			
6	Trust in technology	0.06_b	-0.08_b	0.03_b	0.17_c**	0.08_c	–		
7	Age (years)	-0.01_b	-0.06_b	0.06_b	0.04_c	-0.06_c	0.03	–	
8	Female	-0.08_a	0.08_a	-0.05_a	-0.05_a	-0.02_a	-0.10_b	-0.08_b	–
9	Tertiary education	0.03_a	0.01_a	0.14_a*	0.01_c	0.01_c	0.07_c	0.06_c	0.06_a

*a = Cramer's φc; b = Point biserial correlation r_{pb}; and c = pearson's r. *p < 0.05; **p < 0.01; ***p < 0.001.*

machine forecast do influence the dependent variable. We explore this finding in *post-hoc* analyses after hypothesis testing below.

Participants in the no-test condition assumed that the algorithm-based forecasting model was more accurate ($M = 40.6 \pm 26.4$) compared with participants in the pretest condition ($M = 32.82 \pm 22.3$); Welch's $t(271.65) = 2.64$, $p < 0.01$, $d = 0.32$. Confidence in the algorithm is significantly associated with a higher likelihood of delegating the decision to the algorithm ($\varphi_c = 0.36$, $p < 0.000$), while confidence in the human forecast is negatively correlated with the likelihood of delegating to the algorithm ($\varphi_c = -0.25$, $p < 0.000$). This effect is asymmetric in the sense that higher confidence in the algorithm has a higher positive effect than higher confidence in human forecast has a negative one. Pretesting the algorithm (i.e., seeing it perform and, inevitably, err) has a negative effect on the likelihood of delegating to the algorithm ($\varphi = -0.18$, $p < 0.01$). This is an astonishing finding given that the experiment was designed so that prediction accuracy—i.e., the absolute average prediction error (AAE)—of the human and the algorithmic forecast were virtually identical, with the algorithm's AAE ($M = 20.02 \pm 1.56$) being on average even smaller than the human AAE ($M = 20.71 \pm 5.01$) on the 1–100 performance score; $t(352.4) = -2.19$, $p < 0.05$ (see **Appendix C** for more detailed analyses).

Hypotheses Testing

Figure 2 displays the share (S) of participants who delegated their decision to the algorithm split by condition, vignette, and treatment. We find that participants in the no-test condition and participants in the dismissal vignette scenario were more likely to choose to let the algorithm make the decision. Being able to pretest the quality of the algorithm (see **Figure 2**) [ΔS(*Algorithm pretesting*) = S(no-test) – S(pretest) = 15.1%] had a bigger effect on choice than the type of HR decision [ΔS(*decision type*) = S(Dismissal) – S(Promotion) = 4.4%]. Participants in the *promotion and pretest treatment* ($P_{\_pt}$) were least likely to delegate to the algorithm [$S(P_{\_pt}) = 10.2\%$] and participants in the *dismissal and no-test treatment* ($D_{\_nt}$) were most likely to delegate to the algorithm [$S(D_{\_nt}) = 31.6\%$].

Treatment groups had unequal sizes and were nonnormally distributed but variances were distributed homogenously [Levene's test $F(3, 270) = 2.03$, $p = 0.11$]. **Table 2** presents the odds ratios of delegating to the algorithm by treatment condition. We find no task-related treatment effects purely in relation to dismissing vis-à-vis promoting employees [$\chi^2_{D\_nt/P\_nt}(1) = 0.10$, $p = 0.75$, $\varphi_{D\_nt/P\_nt} = 0.03$; $\chi^2_{D\_pt/P\_pt}(1) = 1.71$, $p = 0.19$, $\varphi_{D\_pt/P\_pt} = 0.11$]. Consequently, H1 finds no support.

However, pretesting the algorithm has a significant effect on the likelihood of delegating the HR decision to the algorithm, irrespective of choice type: Pretesting the algorithm reduces the likelihood of delegating an HR decision to the algorithm [$\chi^2_{nt/pt}(1) = 9.24$, $p < 0.01$, $\varphi_{nt/pt} = -0.18$*]. This effect is stable across both choice task treatments but stronger in the setting of selecting employees for promotion [$\chi^2_{P\_nt/P\_pt}(1) = 6.92$, $p < 0.01$, $\varphi_{P\_nt/P\_pt} = -0.23$*] than for dismissal [$\chi^2_{D\_nt/D\_pt}(1) = 3.18$, $p = 0.08$, $\varphi_{D\_nt/D\_pt} = -0.14$]. However, only the effect of the promotion-related setting is statistically significant and reliable ($\varphi_{P\_nt/P\_pt} = -0.23$*). We investigate the interaction between the valence type of the HR decision task and pretesting further by conduction logistic

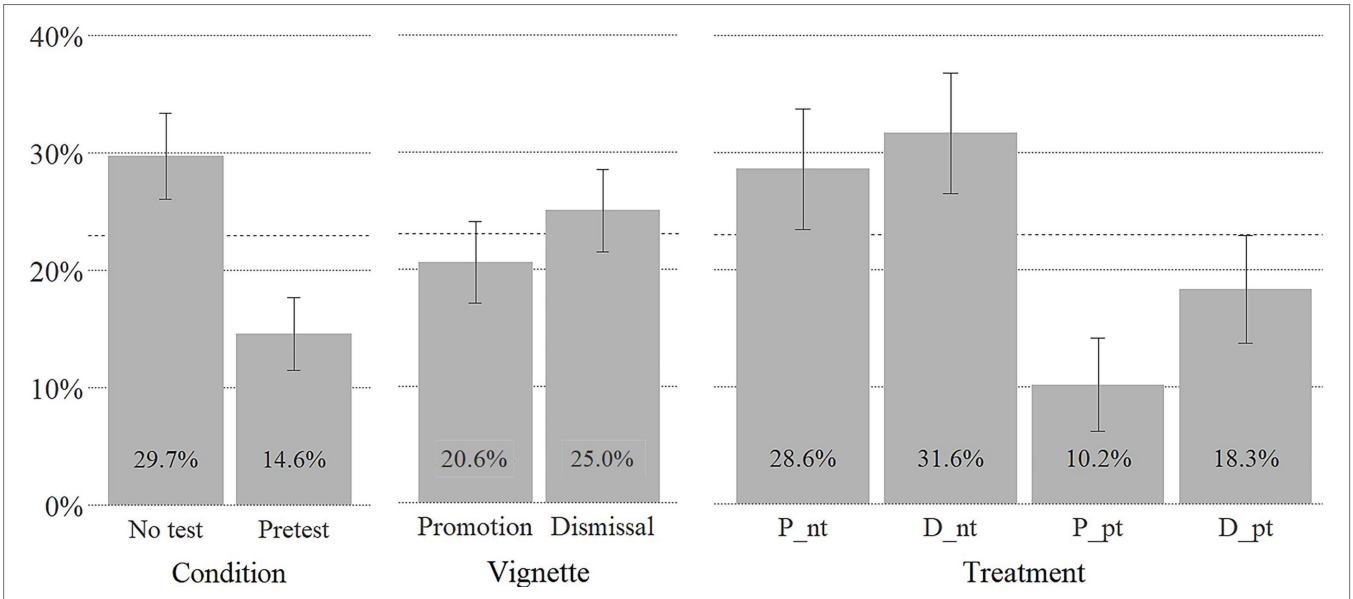

FIGURE 2 | Share (S) of participants delegating human resource (HR) decision to algorithm. Error bars denote ±1 SE; dashed line indicates sample mean of 23%.

TABLE 2 | Odds ratios, χ^2-, and φ-tests of choice by condition, type, and treatment.

	Condition		**HR choice type**		**Treatment**			
	No-test	**Pretest**	**Promotion**	**Dismissal**	**P_nt**	**D_nt**	**P_pt**	**D_pt**
n	158	130	136	152	77	81	59	71
n_{Model}	47	19	28	38	22	25	6	13
n_{Human}	111	111	108	114	55	56	53	58
$Odds_{Model}$	0.42	0.17	0.26	0.33	0.40	0.45	0.11	0.22
$Odds_{Human}$	2.36	5.84	3.86	3.00	2.50	2.24	8.83	4.46
$\Delta Odds_{Model}$ [95% CI]	$\Delta Odds_{pt/nt} = 0.41$ [0.21; 0.76]		$\Delta Odds_{D/P} = 1.28$ [0.71; 2.34]			$\Delta Odds_{P\_pt/P\_nt} = 0.29$ [0.09; 0.80] $\Delta Odds_{D\_pt/D\_nt} = 0.5$ [0.21; 1.14] $\Delta Odds_{D\_nt/P\_nt} = 1.12$ [0.53; 2.34] $\Delta Odds_{D\_pt/P\_pt} = 1.97$ [0.64; 6.79]		
χ^2-tests	$\chi^2_{pt/pt}(1) = 9.24$, $p < 0.01$		$\chi^2_{D/P}(1) = 0.79$, $p = 0.37$			$\chi^2_{P\_nt/P\_pt}(1) = 6.92$, $p < 0.01$ $\chi^2_{D\_nt/D\_pt}(1) = 3.18$, $p = 0.08$ $\chi^2_{D\_nt/P\_nt}(1) = 0.10$, $p = 0.75$ $\chi^2_{D\_pt/P\_pt}(1) = 1.71$, $p = 0.19$		
φ	$\varphi_{nt/pt} = -0.18$*		$\varphi_{D/P} = 0.05$			$\varphi_{P\_nt/P\_pt} = -0.23$* $\varphi_{D\_nt/D\_pt} = -0.14$ $\varphi_{D\_nt/P\_nt} = 0.03$ $\varphi_{D\_pt/P\_pt} = 0.11$		

*$p < 0.05$.

regression analyses (see Model III in **Table 3**) and find no statistically significant interaction between the promotion and dismissal choice context (odds ratio: 0.609, $p = 0.502$). This means that pretesting the algorithm reduces decision makers' likelihood of delegating to the algorithm, irrespective of the choice contest, supporting the baseline hypothesis H2a but not H2b.

Explorative Analyses

Since the effect of algorithm aversion is so prevalent in our data, we conduct *post-hoc* analysis to further investigate the role of CIH vis-à-vis machine-based (CIM) forecasting on individuals' likelihood of delegating HR decisions to an algorithm. Correlation analysis (**Table 1**) revealed that both forms of confidence affect the likelihood of delegation: Higher confidence in machine-based forecasting is correlated with a higher likelihood of delegating to the algorithm ($\varphi_c = 0.36$, $p < 0.000$), and pretesting it reduces the confidence in machine-based forecast ($\varphi_c = -0.18$, $p < 0.000$). Similarly, higher confidence in human forecasting is significantly correlated with a lower likelihood of delegating an HR decision to an algorithm ($\varphi_c = -0.25$, $p < 0.000$).

Logistic regression was used to further analyze the relationship between pretesting the algorithm, choice frame, CIM, and CIH

TABLE 3 | Logistic regression results on choice to delegate HR decision to algorithm.

	Model I				Model II				Model III			
	Odds ratio	**SE**	**[95% CI]**		**Odds ratio**	**SE**	**[95% CI]**		**Odds ratio**	**SE**	**[95% CI]**	
Treatment effects												
Pretesting the algorithm	0.448*	0.159	0.223	0.900					0.546	0.250	0.223	1.339
HR choice type: dismissal	0.589	0.211	0.291	1.190					0.703	0.311	0.295	1.675
Combined treatment effects												
Promotion and pretest (P_pt)						*– reference category –*						
Dismissal and pretest (D_pt)					2.336	1.410	0.716	7.624				
Dismissal and no pretest (D_nt)					3.007†	1.742	0.966	9.357				
Promotion and no pretest (P_nt)					4.278*	2.472	1.378	13.280				
Interaction effects												
Pretesting × dismissal									0.609	0.450	0.143	2.588
Control variables												
Confidence in algorithm forecast	2.483***	0.490	1.687	3.655	2.478***	0.490	1.683	3.650	2.478***	0.490	1.683	3.645
Confidence in human forecast	0.447***	0.099	0.290	0.690	0.451***	0.100	0.292	0.697	0.451***	0.100	0.292	0.697
Trust in technology	0.969	0.326	0.501	1.873	0.993	0.336	0.511	1.927	0.993	0.336	0.511	1.927
Age	0.993	0.029	0.938	1.052	0.994	0.029	0.939	1.052	0.994	0.029	0.939	1.052
Female	0.916	0.326	0.456	1.841	0.927	0.331	0.460	1.866	0.927	0.331	0.460	1.867
Higher education	1.108	0.423	0.524	2.342	1.105	0.422	0.523	2.335	1.105	0.422	0.523	2.335
Constant	0.511	0.677	0.038	6.862	0.099†	0.135	0.007	1.414	0.425	0.576	0.030	6.053
N	267				267				267			
LR χ^2(df)	55.29***				55.75***				55.75***			
df	8				9				9			
Pseudo-R^2	0.200				0.201				0.201			
Log likelihood	−110.81				−110.58				−110.58			

Post-hoc analyses. †$p < 0.10$. *$p < 0.05$; ***$p < 0.001$.

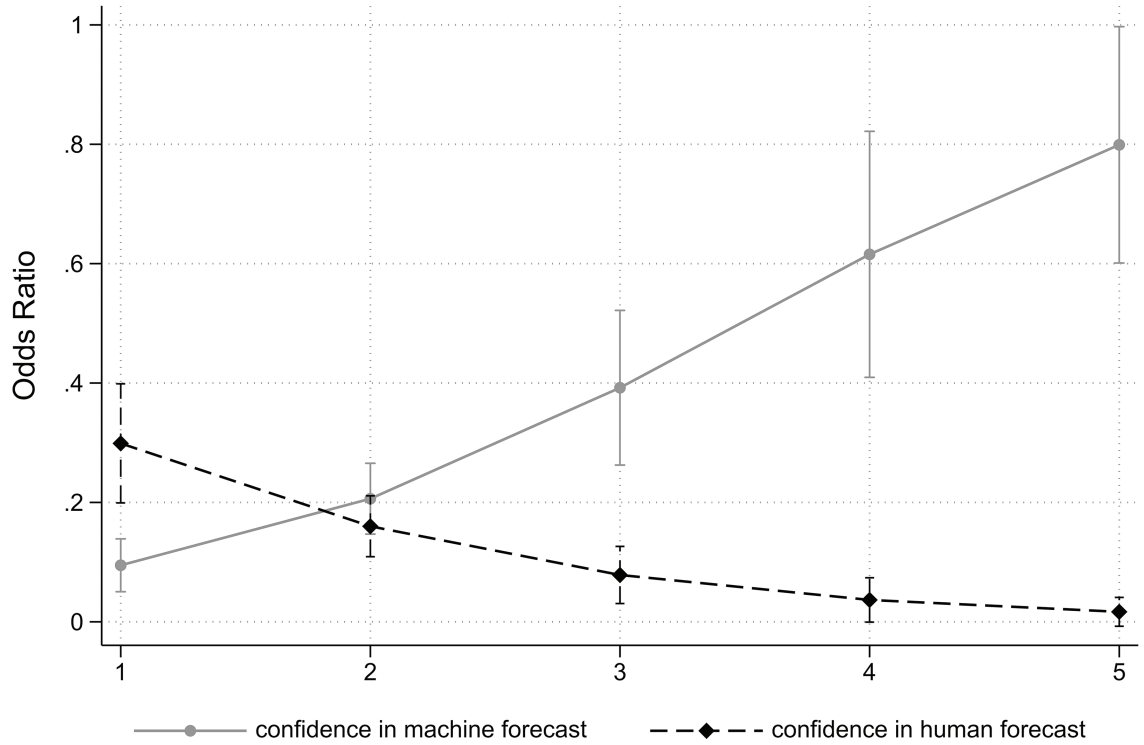

FIGURE 3 | Marginal effects plot of confidence in human (CIH) and machine (CIM) forecast on choice to delegate HR decision to algorithm.

on the probability of delegating the HR decision to the algorithm (see **Table 3**). As displayed in Model I of **Table 3**, we find that, holding the other variables constant, the odds of delegating to the algorithm decreased by 55.2% (95% CI [0.223, 0.900]; $p = 0.024$) for study participants who pretested the algorithm. In the dismissal choice frame, the odds of delegating to the algorithm decreased by 41.1% (95% CI [0.291, 1.190]) but this effect is not statistically significant ($p = 0.140$), supporting the findings presented in the previous section. Furthermore, confidence in the machine forecast (CIM) dramatically increases the odds of delegating to the algorithm (odds ratio: 2.483, $p < 0.000$). For each marginal increase on the five-point CIM-scale, individuals were 148% (95% CI [1.687, 3.655]) more likely to delegate. While confidence in human forecast (CIH) also significantly influences the choice to delegate, its effect is smaller. For each marginal increase in CIH, the likelihood of delegating to the algorithm decreased by 55.3% (95% CI [0.290, 0.690]), these divergent marginal effects are illustrated in **Figure 3**. In Model II, we investigate the relation between the combined treatment effects (HR choice type and pretest condition) but the effects of the confidence variables CIH and CIM remain equally strong. Further analysis revealed no other significant interaction effects.[3]

Since individuals may hold divergent levels of confidence in human and machine forecast, we illustrate their effect on the choice to delegate the HR decision to the algorithm further in **Figure 4**, which displays the share of respondents delegating their HR decision to the algorithm by treatment and clustered by individual confidence levels to account for individual differences in confidence configurations. Individuals who are more confident in human vis-à-vis algorithm-based decision making are less likely to delegate their decisions (0.0–17.6% of respondents). While variation within this group exists in relation to the type of choice task (dismissing vis-à-vis promoting employees) and pretesting the algorithm (test vis-à-vis no-test), the differences are not statistically significant. Among respondents equally confident in model and human forecast, only 14.3% delegate dismissal decisions to the algorithm if they had the chance to pretest it, but 33.3% do so if they did not pretest the algorithm. We find no equivalent discrepancy for the promotion scenario (P_pt = 18.8%; P_nt = 14.3%). Individuals more confident in machine-based forecasts exhibit the strongest effects. Of this group of respondents, 68.4% will delegate dismissal decisions to an algorithm, but only 46.7% do so if they had a chance to pretest the algorithm. Similarly, 48.0% individuals with relatively higher confidence in machine-based decision making will delegate promotion decisions to the algorithm, but only 23.1% do so after testing the algorithm.

DISCUSSION

The experimental findings reveal that changing the valence type of HR decision from a presumably unpleasant decision (i.e., dismissing staff) to a presumably preferable one (i.e., promoting

[3]Additional results of interaction effects analyses are available from the authors upon request.

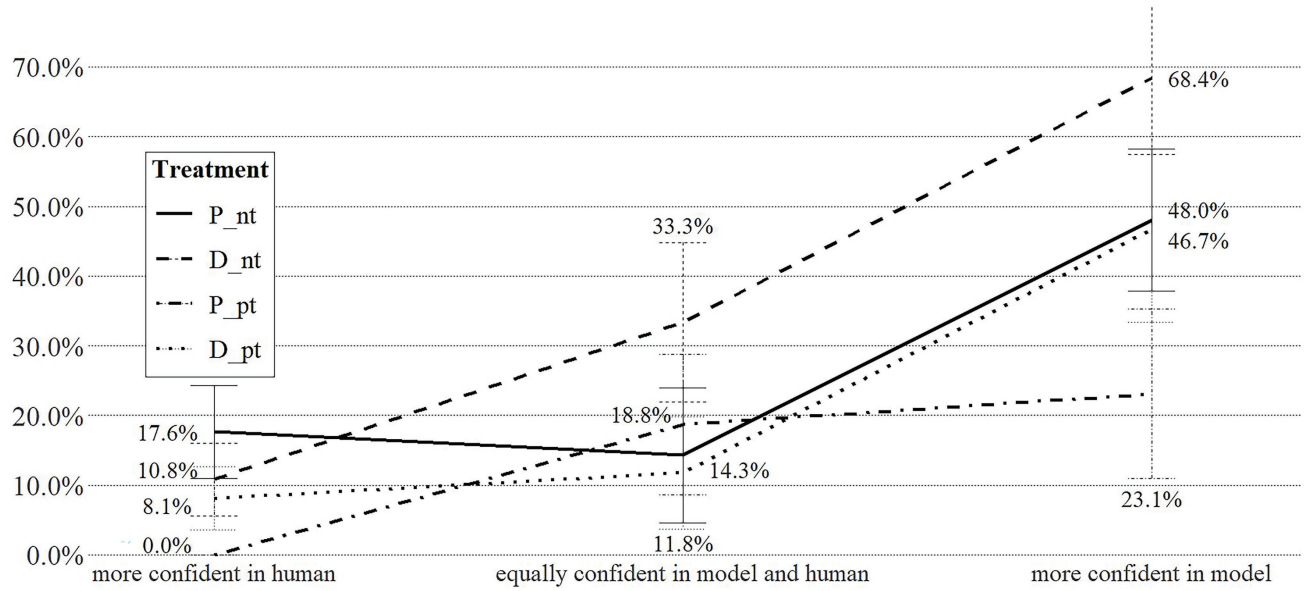

FIGURE 4 | Share (S) of participants delegating HR decision to algorithm, by treatment and confidence levels.

staff) does not necessarily affect the likelihood of delegating this choice to a supportive algorithm. The absence of a substantial direct effect in a strictly controlled experimental study with a relevant sample of young professionals and graduate students has important implications for human resource management scholarship and practice: It shows that the affective reference frame of an HR decision does not function unconditionally as a reliable predictor for whether decision makers will use algorithmic decision support systems to avoid blame and shift the mental burden of responsibility. While there is some indicative support for a choice type-related effect, this effect is much weaker than anticipated and conditional upon the perceived quality of the algorithm forecast as well as decision makers' confidence in machine vis-à-vis human forecast quality. People will not automatically use the algorithm to shift blame.

This is the first experimental study to investigate algorithm aversion and blame avoidance in HR using a German sample. We find the same direction of effects as Dietvorst et al. (2015, 2018) and have hence replicated and extended their results. The study of Dietvorst et al. (2015) was conducted with MBA students from a United States university, while our study relies on experimental data raised with a sample of both young professionals and graduate students from Germany. We contribute to the generalizability of the findings regarding algorithm aversion internationally and in practice, complementing recent scholarship by, among others, Lee (2018), Newman et al. (2020), Filiz et al. (2021), and Renier et al. (2021). Although our findings on algorithm aversion are statistically significant and robust, the effect sizes we observe are substantially smaller—i.e., only one third as large as in Dietvorst et al. (2015). One explanation for the differences in effect sizes is that country cultures exhibit specific differences in trust in artificial intelligence (Gillespie et al., 2021).

The experimental evidence of our study relies on an innovative, balanced, and randomly controlled experiment to warrant high internal validity and to eradicate the influence of socio-demographic factors, which might differentiate employees in their interaction with technology and prime their decision delegation likelihood (McKnight et al., 2011). Neither respondents' trust in technology, age, gender, nor level of education, explained any substantial amount of variance in delegating choice. This is in line with study of Dietvorst et al. (2015), which did not find significant effects regarding these control variables either. This similarity underlines that our results are substantial regarding the observation that blame avoiding behavior is conditional while algorithm aversion is a fundamental psychological mechanism.

Compared with participants in the pretest condition, participants in the no-test condition assumed that the algorithm-based forecasting model was more accurate. This means that testing—and thereby experiencing the fallibility of an algorithm—increases algorithm aversion rather than decreasing it. This finding is in line with prior empirical findings of Dietvorst et al. (2015, 2018), Burton et al. (2020), and Prahl and van Swol (2021), and it highlights practitioners' peculiar challenges in encouraging the use of algorithm-based decision support systems in reluctant staff.

Supporting prior research on blame avoidance in other fields of decision research (Vis and van Kersbergen, 2007; Bartling and Fischbacher, 2012), we specifically contribute to HRM scholarship by revealing that study participants tasked with dismissing staff tend to delegate to the algorithm but only under certain conditions related to their confidence in human and machine forecast, echoing prior findings on the essential role of confidence in human and machine forecast by Filiz et al. (2021). In contrast, pretesting the algorithm reduces the likelihood

of delegating an HR decision to it. This effect is stable across both types of HR decisions assessed, but it is stronger in the setting of selecting employees for promotion than for dismissal, all else being equal. However, only the effect of the promotion-related setting is statistically reliable. This implies that decision makers tend to be less likely to delegate presumably more pleasant promotion decisions to an algorithm, but this relationship is only statistically reliable if pretesting is possible. One explanation for this pattern is that choice delegation may indeed come with a loss of psychological ownership, which is affectively undesirable for (presumably) more pleasant tasks that involve hedonic utility for the decision maker (Önkal et al., 2009; Cassotti et al., 2012; Stark et al., 2017). It is, therefore, individually rational not to delegate if individuals perceive their decision to result in affectively pleasant outcomes in social contexts (Forgas, 2006). Another explanation relates to the flawed expectation of ultimate perfection in algorithms' precision and performance when used in algorithm-based decision support systems in HR (Boyd and Crawford, 2012; Ziewitz, 2016). Individuals mostly expect algorithms to be infallible (Dietvorst et al., 2015; Dietvorst and Bharti, 2020; Berger et al., 2021). When confronted with algorithms' realistic error margins, this expectation disconfirmation may trigger the psychological effect of dissatisfaction-based negativity bias (Oliver, 1980; Oliver and DeSarbo, 1988). This means that experiencing the potentially unexpected fallibility of an algorithm triggers an asymmetrically larger negative response than experiencing the strengths of the algorithm triggers a positive response, even though the algorithm still outperforms human forecast precision. Since recent research by Renier et al. (2021) support this presumption of algorithm-error related negativity bias, we assume that the widely held expectation of algorithm-based HR decision support systems as perfect (Constantiou and Kallinikos, 2015; Leicht-Deobald et al., 2019) may have caused respondents to experience the realistic imperfection of the algorithm in our experiment as a negative surprise, which may have increased reluctance to use it (Madhavan and Wiegmann, 2007; Prahl and van Swol, 2017; Lee, 2018).

We find that decision makers' level of confidence in human vis-à-vis machine-based forecast is a strong predictor of their likelihood of delegation, which is in line with recent empirical findings by Berger et al. (2021). Higher confidence in machine-based forecasting is correlated with a higher likelihood of using the decision support system, which poses a paradoxical practical challenge because pretesting an "imperfect" (i.e., realistic) algorithm *reduces* the confidence in machine-based forecasting. This corresponds to prior findings by Lee (2018). Likewise, higher confidence in human forecasting is significantly correlated with a lower likelihood of delegating HR decisions to an algorithm. Individuals who are more confident in human vis-à-vis algorithm-based forecast precision have a very low likelihood of delegating the decision. Confidence in the algorithm is significantly associated with a higher likelihood of delegating the decision to the algorithm, whereas confidence in the human forecast is negatively correlated with the likelihood of delegating to the algorithm. This effect is asymmetric in the sense that higher confidence in the algorithm has a higher positive effect than higher confidence in human forecast has a negative one. This is in line with prior research on human–human vis-à-vis human–machine trust (Madhavan and Wiegmann, 2007; Prahl and van Swol, 2017, 2021; Filiz et al., 2021). Practitioners need to be aware that familiarity with an algorithm—i.e., the chance to pretest it—may lead to asymmetries in the likelihood of using the algorithm despite the algorithm's usefulness and high forecast precision. Algorithm aversion is, under some conditions, choice task dependent (see also Castelo et al., 2019), but the nature of the choice task is not the decisive factor. While presumably more pleasant HR decisions such as promoting employees may reduce the likelihood of using algorithmic support it is important to note that the combination of pretesting, low confidence in machine forecast, and task type may lead to biased choices and unintended outcomes. Practitioners are encouraged to help their staff build organic trust in their decision support systems but also foster awareness of both the advantages but also risks in using algorithm-based help in HRM (Leicht-Deobald et al., 2019).

Limitations and Future Research

As all empirical research, the generalizability of the findings of the current experiment is limited to some extent. First, while this study relies on data of a highly educated sample of individuals at the start of their careers, it is a convenience sample and not representative for the general population. Yet, in view of the aim of the study, the sample is relevant as it resembles the typical socio-demographic profile of university graduates that are highly in demand for diverse types of managerial training positions in Germany. Furthermore, we are confident in the generalizability of our results (within certain boundaries) since our study replicates and extends the results by Dietvorst et al. (2015).

We identify two specific avenues for future research. First, we believe that more experimental research—especially using realistic treatment conditions with an elevated level of respondent immersion—is needed to further explore the effect of perceived algorithm forecast accuracy as a necessary condition for blame avoidance. Our findings suggest that the likelihood of using an algorithm to support their HR decisions is contingent upon the type of decision but also confidence in, expectations toward, and prior interaction experience with algorithmic forecasting models. Future studies are encouraged to replicate and enhance our experimental design by systematically manipulating the quality of the algorithm to investigate whether the effect is linear or dynamic. Particularly, we encourage future research to replicate our research design with a specific focus on HR professionals' and managers' perspective to explore the effect of professional experience and confidence in human vis-à-vis machine-based forecasts in more detail.

Second, more research is needed to explore the phenomenon of the saturation effect of algorithm aversion. Our data reveal the same stable saturation level as Dietvorst et al. (2015), while using a sample from another country, i.e., Germany. Although this strongly indicates that there is a base-line psychological mechanism at work, more quantitative replication studies with samples from other countries, cultures, and HR tasks are needed to assess its generalizability.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

Open access funding was provided by the University of Bern.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.779028/full#supplementary-material

REFERENCES

Aguinis, H., and Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organ. Res. Methods* 17, 351–371. doi: 10.1177/1094428114547952

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., and Stuart, M. (2016). HR and analytics: why HR is set to fail the big data challenge. *Hum. Resour. Manag. J.* 26, 1–11. doi: 10.1111/1748-8583.12090

Atzmüller, C., and Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology* 6, 128–138. doi: 10.1027/1614-2241/a000014

Bartling, B., and Fischbacher, U. (2012). Shifting the blame: on delegation and responsibility. *Rev. Econ. Stud.* 79, 67–87. doi: 10.1093/restud/rdr023

Bem, D. J. (1972). Self-perception theory. *Adv. Exp. Soc. Psychol.* 6, 1–62. doi: 10.1016/s0065-2601(08)60024-6

Berger, B., Adam, M., Rühr, A., and Benlian, A. (2021). Watch me improve—algorithm aversion and demonstrating the ability to learn. *Bus. Inf. Syst. Eng.* 63, 55–68. doi: 10.1007/s12599-020-00678-5

Boudreau, J., and Rice, S. (2015). Bright, shiny objects and the future of HR. how juniper networks tests and integrates the most valuable new approaches. *Harv. Bus. Rev.* 72–78.

Boyd, D., and Crawford, K. (2012). Critical questions for big data. *Inf. Commun. Soc.* 15, 662–679. doi: 10.1080/1369118X.2012.678878

Burton, J. W., Stein, M.-K., and Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *J. Behav. Decis. Mak.* 33, 220–239. doi: 10.1002/bdm.2155

Cassotti, M., Habib, M., Poirel, N., Aïte, A., Houdé, O., and Moutier, S. (2012). Positive emotional context eliminates the framing effect in decision-making. *Emotion* 12, 926–931. doi: 10.1037/a0026788

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

Constantiou, I. D., and Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. *J. Inf. Technol.* 30, 44–57. doi: 10.1057/jit.2014.17

Dietvorst, B. J., and Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychol. Sci.* 31, 1302–1314. doi: 10.1177/0956797620948841

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *J. Exp. Psychol. General* 144, 114–126. doi: 10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2018). Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them. *Manag. Sci.* 64, 1155–1170. doi: 10.1287/mnsc.2016.2643

Dzindolet, M. T., Pierce, L. G., Beck, H. P., and Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Hum. Factors* 44, 79–94. doi: 10.1518/0018720024494856

Erat, S. (2013). Avoiding lying: the case of delegated deception. *J. Econ. Behav. Organ.* 93, 273–278. doi: 10.1016/j.jebo.2013.03.035

Fildes, R., and Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces* 37, 570–576. doi: 10.1287/inte.1070.0309

Fildes, R., and Hastings, R. (1994). The organization and improvement of market forecasting. *J. Oper. Res. Soc.* 45, 1–16. doi: 10.1057/jors.1994.1

Filiz, I., Judek, J. R., Lorenz, M., and Spiwoks, M. (2021). Reducing algorithm aversion through experience. *J. Behav. Exp. Financ.* 31:100524. doi: 10.1016/j.jbef.2021.100524

Fineman, D. R., Chakrabarti, M., Demetriou, S., Guszcza, J., Houston, J., and Smeyers, L. (2017). "People analytics: recalculating the route: 2017 deloitte global human capital trends." Available at: https://www2.deloitte.com/us/en/insights/focus/human-capital-trends/2017/people-analytics-in-hr.html/#endnote-sup-3 (Accessed December 10, 2021).

Fisher, R. J., and Katz, J. E. (2000). Social-desirability bias and the validity of self-reported values. *Psychol. Mark.* 17, 105–120. doi: 10.1002/(SICI)1520-6793(200002)17:2<105::AID-MAR3>3.0.CO;2-9

Forgas, J. P. (ed.) (2006). *Affect in Social Thinking and Behavior*. New York: Psychology Press.

Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., and Redzepi, A. (2021). The dark sides of people analytics: reviewing the perils for organisations and employees. *Eur. J. Inf. Syst.*, 1–26. doi: 10.1080/0960085X.2021.1927213

Gillespie, N., Lockey, S., and Curtis, C. (2021). *Trust in Artificial Intelligence: A Five Country Study*. Brisbane, Australia: The University of Queensland and KPMG.

Grove, W. M., and Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical–statistical controversy. *Psychol. Public Policy Law* 2, 293–323. doi: 10.1037/1076-8971.2.2.293

Hamman, J. R., Loewenstein, G., and Weber, R. A. (2010). Self-interest through delegation: an additional rationale for the principal-agent relationship. *Am. Econ. Rev.* 100, 1826–1846. doi: 10.1257/aer.100.4.1826

Haube, J. (2015). "HR analytics: a look inside Walmart's HR 'test & learn' model." Available at: https://community.hrdaily.com.au/profiles/blogs/hr-analytics-a-look-inside-walmart-s-hr-test-learn-model (Accessed December 10, 2021).

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Ind. Organ. Psychol.* 1, 333–342. doi: 10.1111/j.1754-9434.2008.00058.x

Hill, A. (2015). Does delegation undermine accountability? Experimental evidence on the relationship between blame shifting and control. *J. Empir. Leg. Stud.* 12, 311–339. doi: 10.1111/jels.12074

Hsee, C. K. (1996). Elastic justification: how unjustifiable factors influence judgments. *Organ. Behav. Hum. Decis. Process.* 66, 122–129. doi: 10.1006/obhd.1996.0043

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Acad. Manag. J.* 38, 635–672. doi: 10.2307/256741

Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291. doi: 10.2307/1914185

Kuncel, N. R., Klieger, D. M., Connelly, B. S., and Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: a meta-analysis. *J. Appl. Psychol.* 98, 1060–1072. doi: 10.1037/a0034156

Langer, M., and 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. doi: 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.* 5:205395171875668. doi: 10.1177/2053951718756684

Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., et al. (2019). The challenges of algorithm-based HR decision-making for personal integrity. *J. Bus. Ethics* 160, 377–392. doi: 10.1007/s10551-019-04204-w

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

Madhavan, P., and Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: an integrative review. *Theor. Issues Ergon. Sci.* 8, 277–301. doi: 10.1080/14639220500337708

McKnight, D. H., Carter, M., Thatcher, J. B., and Clay, P. F. (2011). Trust in a specific technology. *ACM Trans. Manag. Inf. Syst.* 2, 1–25. doi: 10.1145/1985347.1985353

Meehl, P. E. (1954). *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis: University of Minnesota Press.

Mentzer, J. T., and Kahn, K. B. (1995). Forecasting technique familiarity, satisfaction, usage, and application. *J. Forecast.* 14, 465–476. doi: 10.1002/for.3980140506

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

Oexl, R., and Grossman, Z. J. (2013). Shifting the blame to a powerless intermediary. *Exp. Econ.* 16, 306–312. doi: 10.1007/s10683-012-9335-7

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *J. Mark. Res.* 17, 460–469. doi: 10.1177/002224378001700405

Oliver, R. L., and DeSarbo, W. S. (1988). Response determinants in satisfaction judgments. *J. Consum. Res.* 14:495. doi: 10.1086/209131

Önkal, D., and Gönül, M. S. (2005). Judgmental adjustment. A challenge for providers and users of forecasts. *Foresight Int. J. Appl. Forecast.* 1, 13–17.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., and Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *J. Behav. Decis. Mak.* 22, 390–409. doi: 10.1002/bdm.637

Peck, D. (2013). "They're watching you at work." Available at: https://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/ (Accessed December 10, 2021).

Peeters, T., Paauwe, J., and van de Voorde, K. (2020). People analytics effectiveness: developing a framework. *J. Organ. Effective. People Perform.* 7, 203–219. doi: 10.1108/JOEPP-04-2020-0071

Petropoulos, F., Fildes, R., and Goodwin, P. (2016). Do 'big losses' in judgmental adjustments to statistical forecasts affect experts' behaviour? *Eur. J. Oper. Res.* 249, 842–852. doi: 10.1016/j.ejor.2015.06.002

Prahl, A., and van Swol, L. (2017). Understanding algorithm aversion: when is advice from automation discounted? *J. Forecast.* 36, 691–702. doi: 10.1002/for.2464

Prahl, A., and van Swol, L. (2021). Out with the humans, in with the machines? Investigating the behavioral and psychological effects of replacing human advisors with a machine. *Hum. Mach. Commun.* 2, 209–234. doi: 10.30658/hmc.2.11

Reindl, C. U. (2016). People analytics: datengestützte mitarbeiterführung als chance für die organisationspsychologie. *Gruppe Interakt. Organ. Z. Angewandte Organ.* 47, 193–197. doi: 10.1007/s11612-016-0325-7

Renier, L. A., Mast, M. S., and Bekbergenova, A. (2021). To err is human, not algorithmic—robust reactions to erring algorithms. *Comput. Hum. Behav.* 124:106879. doi: 10.1016/j.chb.2021.106879

Rozin, P., and Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personal. Soc. Psychol. Rev.* 5, 296–320. doi: 10.1207/S15327957PSPR0504_2

Sanders, N. R., and Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega* 31, 511–522. doi: 10.1016/j.omega.2003.08.007

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychol. Bull.* 66, 178–200. doi: 10.1037/h0023624

Sharma, A., and Sharma, T. (2017). HR analytics and performance appraisal system. *Manag. Res. Rev.* 40, 684–697. doi: 10.1108/MRR-04-2016-0084

Shrivastava, S., Nagdev, K., and Rajesh, A. (2018). Redefining HR using people analytics: the case of Google. *Hum. Resour. Manag. Int. Dig.* 26, 3–6. doi: 10.1108/HRMID-06-2017-0112

Shteingart, H., Neiman, T., and Loewenstein, Y. (2013). The role of first impression in operant learning. *J. Exp. Psychol. General* 142, 476–488. doi: 10.1037/a0029550

Staab, P., and Geschke, S.-C. (2020). *Ratings als Arbeitspolitisches Konfliktfeld: Das Beispiel Zalando*. Düsseldorf: Hans-Böckler-Stiftung.

Stangor, C., and McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: a review of the social and social developmental literatures. *Psychol. Bull.* 111, 42–61. doi: 10.1037/0033-2909.111.1.42

Stark, E., Baldwin, A. S., Hertel, A. W., and Rothman, A. J. (2017). Understanding the framing effect: do affective responses to decision options mediate the influence of frame on choice? *J. Risk Res.* 20, 1585–1597. doi: 10.1080/13669877.2016.1200654

Steffel, M., Williams, E. F., and Perrmann-Graham, J. (2016). Passing the buck: delegating choices to others to avoid responsibility and blame. *Organ. Behav. Hum. Decis. Process.* 135, 32–44. doi: 10.1016/j.obhdp.2016.04.006

Tambe, P., Cappelli, P., and Yakubovich, V. (2019). Artificial intelligence in human resources management: challenges and a path forward. *Calif. Manag. Rev.* 61, 15–42. doi: 10.1177/0008125619867910

Tversky, A., and Kahneman, D. (1991). Loss aversion in riskless choice: a reference-dependent model. *Quart. J. Econ.* 106, 1039–1061. doi: 10.2307/2937956

Vis, B., and van Kersbergen, K. (2007). Why and how do political actors pursue risky reforms? *J. Theor. Polit.* 19, 153–172. doi: 10.1177/0951629807074268

Vrieze, S. I., and Grove, W. M. (2009). Survey on the use of clinical and mechanical prediction methods in clinical psychology. *Prof. Psychol. Res. Pract.* 40, 525–531. doi: 10.1037/a0014693

Ziewitz, M. (2016). Governing algorithms. *Sci. Technol. Hum. Values* 41, 3–16. doi: 10.1177/0162243915608948

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.