# Hidden Costs of Automation: The Impact of Process Automation on Performance*

***Working Paper***

Christina Strobel†

June 7, 2023

The objective of this paper is to investigate the impact of process automation on human performance, with a specific emphasis on the influence of fairness, trust, and expectations. I also examine the importance of the decision-maker in the process of automation. As a workhorse I use a modified two-stage Principal-Agent Game. Principals can set their bonus payment expectations either with or without knowledge of the actual performance (ex post or ex ante), revealing an automated system's inflexibility. I do not observe any variation in perceived fairness or trust between the two processes, but I find divergent expectations and a significant lower performance if an automated process is used. Yet, the decrease in performance cannot be primarily attributed to fairness and trust issues, but rather to divergent expectations. I also do not observe any difference in the outcomes based on who makes the decision to automate. My findings underscore the importance of expectations within an automated process and their consequential effect on performance, highlighting that automation can also impact the fundamental parameters that underlie the process.

*Keywords*: Automation; Performance; Management Process Automation; Online Experiment; Principal-Agent-Game
*JEL classification*: C91, D63, D80

*This document was created on June 7, 2023, with R version 4.2.2 (2022-10-31 ucrt), on x86_64-w64-mingw32.

†Hamburg University of Technology, Institute for Digital Economics, Am Schwarzenberg-Campus 4 (D), 21073 Hamburg, Christina.Strobel@tuhh.de.

# 1. Introduction

Algorithm-controlled processes are taking over more and more tasks from humans. This is not only the case in manufacturing, where robots are replacing assembly line workers, but also in management, where the level of process automation continues to increase (Aharoni and Fridlund, 2007; Brynjolfsson and Mcafee, 2017; Lee, 2018). Forecasts even predict a compounded annual growth rate of 10.5% for the business process management industry until 2025 mainly driven by advanced business process management (BPM) solutions (Markets and Markets, 2020). The most prevalent examples of this development are online companies such as Uber, Gorillas, and Delivery Hero, which already rely on process automation to a large extent.

A management area in which automated processes are already widespread as of today is human resources (HR).[1] According to a study by Gartner, 70% of HR functions in international companies already used AI-based automation in 2019, and it was predicted that this number will have increased by 30% by 2022 (Berger-Böcker, 2019). A typical task that is automated is assisting in the selection of prospective employees by evaluating and selecting the most suitable candidates (Upadhyay and Khandelwal, 2018; Liem et al., 2018; Van Esch et al., 2019). Other examples are the use of automated processes to optimize incentive compensation schemes (Riberolles, 2020), evaluate performance (Kaur and Sood, 2017; Greenfield, 2018), decide about a pay raise (Fisher, 2019), and execute performance appraisal decisions with minimal human involvement (Yen et al., 2017). Automation is especially important in personnel decisions such as hiring, firing, promotions, and bonuses, as these decisions directly affect another person and include not only a monetary but also a non monetary component.[2]

Automation might increase equality, e.g., reduce personal (potentially biased) impressions (Lawler and Elliot, 1996; Hooper et al., 1998), but simultaneous "HRM [human resource management] algorithms are performative while triggering changes in social processes" (Meijerink et al., 2021, p. 2552). A recent survey on the use of digital human resource management tools, for example, demonstrates that while companies are aware of the benefits of digital human resource management tools, they are also concerned about how they will be perceived by employees (Danilov and Chugunova, 2022). Software solutions based on semantic models are replacing direct human-to-human interactions; thus, human interactions fade from sight (Huang and Rust, 2018). Because of this, decisions affecting the employee depend much less on an individual's assessment than on an ex ante predetermined rules making the process less individualizable and more rigid (Stone, 1971). The automation of tasks that were previously carried out by humans through automated systems can result in processes becoming less flexible, as models and parameters need to be predetermined and they may require substantial effort to modify. Raisch and Krakowski (2021) point out that AI-based solutions create a new tension in management because rule-based automation allows delegation, but complex management tasks can only be delegated to a certain extent, and therefore urge management scholars to conduct research on the use of AI in organizations. However, there

[1] For an extensive overview on algorithmic process automation in HR, see Cheng and Hackett (2021).

[2] For a literature review on the use of AI in tactical workforce management, see Votto et al. (2021).

is very little insight from management scholars into AI and human behavior although it is required as the technology is becoming increasingly important in the management context. Processes are being simplified, quantified and streamlined. This might raise procedural fairness concerns, as it shifts the pivotality for the decision from the human to the algorithm, relieves supervisors from individual decisions, and therefore reduces human agency for the decision (Ötting and Maier, 2018). Automated processes are also often not transparent and offer fewer opportunities to consider exceptional circumstances.

Given the rapid development of new, scalable automated solutions, it is therefore inevitable to investigate how people perceive and respond to automation in social management processes. Not only for situations, where an algorithm functions as an automated colleague, but also for superior-subordinate interactions, where an algorithm manages humans as their superior (Chugunova and Sele, 2022). As circumstances change with the replacement of human interactions by automated processes, it is exceedingly important to find out how performance is affected. The question is: How do people perceive automated processes in management tasks, and does it have an impact on human performance?

The main contribution of this paper is studying the impact of process automation in social management tasks on performance, with a specific focus on procedural fairness, trust, and expectations, while also exploring the significance of the decision-maker (human supervisor or automated third party) in the automation process. I thereby focus on the automation of the bonus payment process as a social management task. Decisions on bonus payments are often made as part of a performance appraisal process. To be eligible for a bonus, a level of performance must first be achieved. If a performance expectation is reached, the employee is entitled to receive the bonus. While in the past a supervisor would have decided on final bonus payments, bonus assessments are now more and more automated, leaving extremely limited room for the supervisor to influence the decision. Thus, the employee's bonus payment depends much less on an individual's assessment than on a predetermined algorithm. Former research indicates that the performance appraisal process matters and that in particular, the execution of the process by the direct supervisor has a positive effect on human motivation and performance (Levy and Williams, 2004; Kampkötter, 2017). This brings up the question of how the automation of the performance appraisal process, making it more rigid and inflexible, will impact performance.

My study contributes to the literature on performance feedback (see Villeval (2020) for a recent survey). While most studies examine worker's performance under different performance evaluations and bonus payments, I focus on the automation of the process, and how automation decisions made by different decision makers are perceived. )) The paper contributes to the literature by studying the link between the use of automation and human work performance in an incentivised experimental setup.

The remainder of the paper is organized as follows: Section 2 provides a literature review focusing on empirical and experimental evidence. In Section 3, I describe the experimental design. In Section 4, I relate the experiment to the theoretical background and derive behavioral predictions. In Section 5, I present the results. Section 7 concludes the paper by summarizing the main findings and discussing their implications as well as identifying further research ideas.

# 2. Related literature

The review focuses on the influence of bonus payments and performance appraisals on performance as well as on research on the determinants of performance, and the automation of (social) management tasks. According to the call for an interdisciplinary effort to research human behavior when interacting with automated agents formulated by Rahwan et al. (2019), the review covers a wide range of findings from different fields, including economics, management, psychology, consumer research, human-computer interaction, and computer science.

## 2.1. Individual performance appraisals and performance

Previous research has shown that, in addition to monetary payments, the process itself has an impact on performance.[3] Using data from the German Socio-Economic Panel (SOEP) Kampkötter (2017) found that a monetary performance appraisal process conducted by the supervisor increases the employee's job satisfaction and leads to higher performance. A literature review of more than 300 management papers on performance appraisal research conducted by Levy and Williams (2004) also indicates that not only the monetary incentive but also the performance process itself matters. The review shows that the supervisor's recognition and appreciation of the work performed are, aside from the monetary benefit, the main factors that cause an increase in an employee's job satisfaction. Along these lines, a recent study by Villeval (2020), on behavioral and experimental research papers, found that the process indeed has a positive cognitive and motivational effect on performance. Principals, however, tend to be lenient on performance feedback because they are uncomfortable delivering bad news to workers. Interestingly, an online experiment by Kusterer and Sliwka (2022) found that supervisors were lenient even when they did not had to expect a negative reaction from workers.

## 2.2. Trust, control, and performance

The effect of bonus payments on performance has been researched extensively and indicates that performance seems not only to be influenced by monetary incentives. The theoretical model by Ellingsen and Johannesson (2005) shows that especially esteem influences performance, as a generous and trusting contract can elicit better performance from agents than a contract with low pay and strong incentives.

The theoretical finding that performance is influenced by situational factors is supported by experimental findings showing that performance depends not only on monetary incentives but also situational factors such as fairness, trust, and control. Fehr et al. (2007), for example, find bonus contracts that rely on fairness and trust as an enforcement device to be more efficient and more profitable than incentive contracts enforced by the courts. In their experiment, the principal was able to choose a mechanism to enforce a specific performance from the agent with the support of a third party or to announce a non binding,

[3] In accordance with the literature on performance appraisal systems, I do not differenciate between performance and effort and both terms are used interchangeably.

voluntary bonus payment instead if the agent's performance was satisfactory. The results show that a non binding, voluntary bonus payment leads to higher performance than an explicit incentive contract, that fines the agent for unsatisfactory performance. Experiments also indicate that employee performance is sensitive to the level of control. Fehr and Rockenbach (2003) and Fehr and List (2004) show that the principal's decision to use a punishment device leads to a decrease in performance by the agent in Trust Games. In a similar spirit, Falk and Kosfeld (2006), Fehr and Schmidt (2007), and Kajackaite and Werner (2015) find that controlling the agent has a negative influence on their performance. In a chosen-effort experiment conducted by Falk and Kosfeld (2006), the agent had to choose a costly performance that benefited the principal while the principal had the choice to either control (i.e., enforce a minimum performance) or trust the agent. The results show that the majority of the agents reduced their performance because most agents perceived control as a signal of distrust and low expectations by the principal. Fehr and Schmidt (2007) find, using a chosen-effort setup in a principal-agent experiment, that bonus payments increase performance and are more efficient than fixed-wage contracts and contracts that include the possibility of punishing the employee for bad performance. Kajackaite and Werner (2015) build upon the finding that control has a counterproductive effect on performance provision by showing that the principal's active decision to control affects the agent's kindness perception and triggers reciprocal responses. However, they find no significant change in the average output level in a real-effort experiment if the principal decides to implement a minimum performance requirement.

The effect of control on performance seems to also be influenced by who is exercising control. Schmelz and Ziegelmeyer (2015) show that the effect of control depends on the closeness between the agent and the principal. By running a Principal-Agent Game on the Internet as well as in a laboratory, Schmelz and Ziegelmeyer (2015) find that exercising control is less likely to reduce work performance in a remote than in a laboratory setting. Control can be exerted not only by the principal but also by a third party. Burdin et al. (2018) found agents show higher performance if principals abstain from control than when a third party decides not to control. In earlier work, however, Falk and Kosfeld (2006) found the opposite: in an experiment based on the Principal-Agent Game, the performance of the agent was higher if control was executed by a third party instead of directly by the principal.

## 2.3. Automated (social) management tasks

Research shows that there is no general aversion to the use of algorithms, but that certain factors such as the performance of the algorithm (Dietvorst et al., 2015) and the ability to override the automated decision (Dietvorst et al., 2018; Bigman and Gray, 2018) have an impact on algorithm aversion.[4] When one turns to decision automation in management tasks, it becomes clear that people seem to accept or even prefer automated managerial decisions for analytical tasks (Cormier et al., 2013; Bai et al., 2022) but not for social tasks (Waytz and Norton, 2014; Lee, 2018; Hertz and Wiese, 2019; Castelo, 2019; Bigman and Gray, 2018).[5] For

[4]For a comprehensive literature review on factors influencing algorithm aversion, see Jussupow et al. (2020).

[5]For extensive literature reviews on the effects of decision automation in human resource management, see Langer and Landers (2021), Meijerink et al. (2021), and Vrontis et al. (2022).

example, Bigman and Gray (2018) find in nine vignette studies that people are averse towards automation in morally-relevant driving, legal, medical, and military decisions. The result seems to be found not only in vignette studies but also in experiments. Hertz and Wiese (2019) find in a non-monetarily incentivized online experiment that algorithms are less likely to be selected to perform a subjective tasks (reading the mind in the eyes task) than an analytical tasks (calculation task) and are less trusted.

Automation also appears to influence how the process is perceived in terms of trust and fairness. The finding that algorithms are perceived as less trustworthy in subjective tasks than in objective tasks was also found in two large vignette studies on Amazon Mechanical Turk (MTurk) conducted by Waytz and Norton (2014) and Castelo (2019), each asking to what extent one would trust an algorithm or a very well qualified person on different subjective or objective tasks. Regarding the perceived fairness of the process, Nagtegaal (2021) finds that participants in a survey experiment rate decisions by managers as higher in procedural justice than automated decisions. However, this was the case for complex tasks, while the opposite was true for tasks with low complexity. However, Wang et al. (2020) find in a vignette experiment on factors influencing the perceived fairness of algorithmic decision-making that people judge the algorithm to be fairer if the algorithm predicts in their favor, suggesting the existence of a self-serving bias.

Automation also has a particularly strong influence on the perception of HR processes. Kaibel et al. (2019) find in two experimental studies a negative effect of algorithm-based staffing decisions on the perceived personableness of the process and the organizaional attractiveness. This may be due to the fact that automated decisions on subjective issues are perceived differently than human decisions. Indeed, by manipulating the superior in a vignette study, Höddinghaus et al. (2021) find that while automated decisions score higher when it comes to integrity and transparency, they are perceived as less adaptable and less benevolent than human decisions in a hiring task. However, the evidence regarding the perceived fairness of an automated process is rather inconclusive. Contrary to the findings that an automated decisions process is perceived to be less fair than human decisions, Ötting and Maier (2018) find in a vignette experiment no interaction effect between procedural justice measurements and the type of decision agent (human, robot, or computer) in a job allocation task.

Turning to automated performance evaluations, research shows that automated processes are also perceived differently than non automated processes. Lee (2018) find in a vignette study, where the decision-maker (algorithmic or human) in four managerial tasks (work assignment, scheduling, hiring and evaluation) was manipulated, that if an algorithm is used to conduct a performance review, the process is judged as less fair and trustworthy. The effect is moderated by the algorithms' perceived lack of intuition and subjective judgment capabilities. While algorithmic decisions were perceived as less fair and trustworthy in tasks that need more subjective judgement and emotional capacities, they are perceived as equally fair and trustworthy in mechanical tasks. Newman et al. (2020) confirm in a vignette study the finding that people judge an automated performance review process to be less fair and document that this might be due to reductionism and procedural fairness concerns. The authors argue that automating the performance review process reduces the qualitative aspects of the performance to quantifiable metrics (i.e., quantification) and thus fails to evaluate

performance in a broader context (i.e., decontextualization). Consequently, decisions by algorithms are perceived to be based on less accurate (i.e., incomplete) information than those made by humans and are thus perceived to be less fair.

In summary, previous research suggests that not only human involvement, but also the nature of the task as well as the rigidity of an automated process and the resulting impact on the perceived fairness of and trust in the process may play a role in the acceptance of automated management processes.

# 3. Experimental design

I implemented an experimental design based on a Principal-Agent Game, an automated and a non automated decision process, two treatment, and a questionnaire to measure the perceived fairness, trust, and expectations.

## 3.1. Automated and non automated decision process

The main goal of the design is to establish a situation that enables the observation of participants' performance in both automated and non-automated processes. In order to closely replicate a management situation, I employ the widely recognized Principal-Agent paradigm, which provides three distinct advantages. First, it effectively emulates a subordinate-supervisor relationship. Second, it enables the representation of a performance appraisal process. Third, it facilitates a standardized comparison of the two processes within a well-known and frequently used game. In modeling the game, I incorporate the two-stage Principal-Agent Game introduced by Falk and Kosfeld (2006), which is outlined as follows: the agent engages in a productive activity which is costly to the agent but beneficial to the principal; the agent has an initial endowment of 120 points (1 point =\$0.01), while the principal's initial endowment is 0 points; the agent, in turns, chooses a productive activity $x$, and the cost of the productive activity for the agent is $c(x) = x$; the principal earns twice the agent's performance $p(x) = 2x$. I expand upon the framework proposed by Falk and Kosfeld (2006) by adding a bonus system, which enables me to encompass automation. In the bonus system, the principal sets a performance threshold $x_t$ that the agent must reach in order to receive a bonus $b^* \in \{0, 120\}$. A binary mechanism is used as it allows the agent to form beliefs about the bonus payment more easily than a complex mechanism. The bonus $b^*$ is paid by the experimenter. The agent receives the bonus if $x_t \geq x$. Thus, the payoff functions are $\prod_P = 2x$ for the principal and $\prod_A = 120 - x + b^*_{x_t}$ for the agent. To design the non automated decision process and the automated decision process, I focus on the fundamental difference between the two processes: process rigidity. What finally distinguishes automated from non automated processes is the fact that features need to be fixed in advance to calibrate the automated process.[6] While people in an non automated decision process possess spontaneity, i.e., the ability to make a decision completely freely and under absolute freedom of choice

[6]With machine learning and neural network approaches, more and more of these features can be defined automatically by algorithms in the process, but it is still necessary to provide certain parameters and data to the algorithm in advance, which the algorithm uses to calibrate itself.

at any time, this must be given up to a certain extent in order to make automation possible in the first place. This ultimately reduces the influence that a person has on a decision. Compared to a non automated process, an automated process is therefore by definition more rigid and less individualizable. I thus focus on the rigidity of both processes wherein the influence of the principal's actual decision-making power over the payment of the bonus is considered as the diverging factor between the non automated and the automated process. To model the rigidity of an automated process, the principal's expectation for bonus attainment is pre-determined without any awareness of the agent's actual performance, while in a non-automated bonus payment process, the expectation is formulated based on knowledge of the agent's performance. More technically speaking, the principal knows the agent's productive activity $x$ before determining the expectation $x_t$ for the agent to get the bonus $b^*$ in the non automated decision process. Thus, the principal decides on the expectation about the agent's performance to quantify for a bonus *ex post*, after knowing the agent's actual performance. The principal thus has absolute power to decide whether a bonus is paid or not, and at the moment of the decision there is also absolute certainty about the impact of the decision on the agent. To model the automated decision process, the principal does not know the agent's productive activity $x$ before determining the expectation $x_t$ for the agent to reach to get the bonus $b^*$. Therefore, the principal determines the expectation for the bonus *ex ante* to the performance. Thus, the principal has only a conditional authority to decide whether or not to pay a bonus and does not know directly at the time of the decision how the decision will affect the agent. The process is therefore less individualizable and more rigid. This is analogous to the process of automation, where decision parameters are predetermined to calibrate an algorithm.

## 3.2. Treatments

Depending on the treatment, either the principal (treatment *HUMAN*) or a automated third party (treatment *SYSTEM*) decides whether to use the non automated or automated decision process. While in treatment *SYSTEM* a random mechanism decides with a probability of 50% to use either the automated or the non automated decision process the decision is incumbent on the principal in treatment *HUMAN*. More precisely, in treatment *HUMAN*, the principal first decides on whether to use a non automated or an automated decision process and then determines the expectation for the bonus payment. Because of this, in the non automated decision process, the principal not only sets an expectation but also determines immediately whether the agent receives a bonus or not, as the principal knows the actual performance when determining the bonus expectation.

## 3.3. Measurement of perceived fairness and trust

After the principals and agents made their choice, but before participants are informed about the final outcome and payoff, agents are asked about the procedural fairness of the process as well as their trust in receiving a bonus. I use these questions as a proxy for the perceived fairness of and trust in the process. As participants can refer to different concepts of fairness, following Konow (1996) a simplified dichotomous scale is used to measure the

perceived overall subjective process fairness. Following the debate over the length of the scale a four-item scale is used to measure trust in receiving a bonus (see Bauer and Freitag, 2018). To account for the possibility that the trust and fairness measurements are mainly driven by diverging expectations in an automated decision process versus a non automated decision process, agents are also asked to state their expectations about receiving the bonus in both processes.[7] Agents also have the option of explaining why they choose a specific performance in the non automated decision process and automated decision process in an open question. Finally, participants are asked about their general risk (Dohmen et al., 2011) and trust preferences (Sturgis and Smith, 2010), their age, and their gender. All questions can be found in Appendix A.8.

## 3.4. Procedure

The experiment was conducted online via MTurk, a website for crowdsourced labor, using oTree (Chen et al., 2016). The participants had to have completed at least 100 so-called human intelligence tasks (HITs) on Amazon MTurk and had to have an approval rate of 99% for their completed HITs, to take part in the experiment. The purpose of these requirements was to ensure the successful execution of the experiment, as a premature departure from the study could potentially affect its feasibility. Furthermore, these requirements enabled a certain level of control over the pool of participants, which, in turn, influenced the quality of the generated data. All experimental stimuli and instructions were presented through a computer interface. Participants received a participation fee of $0.50 in addition to the money they earned during the experiment.

| | |
|---|---|
| Stage 1 | Human test, Instructions, Group matching, Role assignment |
| Stage 2 | Control questions |
| Stage 3 | Agent decides about performance (strategy method)<br>Principal decides about decision process (treatment *HUMAN*) |
| Stage 4 | Principal decides about bonus expectation |
| Stage 5 | Questionnaire |

Table 1: Procedure.

As Table 1 shows, all participants had to pass a test to ensure only humans would participate in the experiment. Therefore, the participants had to add two two-digit numbers and write the correct answer in an input field. Participants who passed the human test were randomly assigned to a group of two as well as to a role and were provided with the experimental instructions.[8] Each group consisted of one agent (participant A) and one principal (participant B). Participants went through five different stages: In stage 1, all participants had to

[7] Agents' expectations about the performance threshold were not incentivized. Not incentivizing expectations also prevents hedging between the decision on how much to transfer and the expected payoff as a result of a correct expectation.

[8] Instructions can be found in Appendix A.7.

pass a human test to ensure that only humans participated in the experiment and received the experiment instructions. In stage 2, all participants had to answer a set of control questions to ensure they understood the instructions before proceeding. In stage 3, depending on the treatment, principals had to make a decision about whether a non automated or automated decision process was going to be used. The order in which both systems were presented was randomly alternated for each participant in both treatments to control for potential order effects. As Falk and Kosfeld (2006) do not indicate a difference between using the strategy method or the specific response method and Charness et al. (2018) find qualitatively similar results for real-effort and stated-effort designs in a meta-study on performance measures in economic experiments, a strategy method was used to elicit the agent's performance. The strategy method also offers the opportunity to examine the effect on performance regardless of the task type; creative or cognitive. Therefore, agents were asked to state their performance in a non automated decision process as well as in an automated decision process. In stage 4, principals had to decide on the bonus expectation that had to be reached to get the bonus either before (automated decision process) or after (non automated decision process) knowing the performance by the agent. In stage 5, the agents and the principals were asked about how the process was perceived, and some descriptives were collected by the help of a short questionnaire.

# 4. Behavioral predictions

Automated decisions are based on standardized processes and built upon predetermined rules that are programmed *ex ante* to the occurrence. Non automated decisions are more situation specific, and hence *ex post* driven. Therefore, automation does not necessarily change who decides but rather when a decision is made. Thus, the agents' expectations regarding the bonus expectation should not change, as the expectation set by the principal should be independent of when the decision was made.[9] What changes, however, is the process by which the expectation is determined. Whereas with a non automated decision the expectation is defined only after the performance, with an automated decision this is done in advance. This changes the way decisions are made. Previous research found that reactions toward automated decisions depend on situational circumstances such as task type, control, transparency, trust, and process fairness (Chugunova and Sele, 2022; Langer and Landers, 2021). While the task in the experiment remains the same, as does the principal's control over the bonus payment and the transparency of the process, automation makes the performance appraisal process more rigid as spontaneity decreases, thus changing the performance appraisal process itself. The decision is no longer a individual decision *ex post* to the performance but rather a rigid *ex ante* benchmark that was set before actually knowing the performance. Therefore, the perceived fairness of the process and the confidence in receiving a bonus may change. People faced with an automated process must therefore decide how they value the process compared to a non automated process.

[9] This is also what is predicted by standard economic theory; that is, certain performances by agents are more likely to occur (see Appendix A.3) but does not predict any difference between an automated and a non automated process.

According to the literature, factors such as situation-specific trust in and perceived fairness of the process play an important role in the evaluation of an automated process. Previous work on procedural justice shows that people perceive a decision process as fairer if the processes are consistent, based on accurate information, and are unbiased (Levanthal, 1980; Thibaut and Walker, 1975). However, experimental research on procedural fairness finds automated decisions to be perceived as more rigid and decontextualized than non automated decisions (e.g., Newman et al., 2020; Ötting and Maier, 2018). Automated decisions might also not present direct means to voice an opinion or correction, lowering the possibility that situational factors can be considered. It also provides the opportunity to relieve supervisors from the imminent decision. This could lead to both, a decrease and an increase in procedural justice and trust compared to non automated decisions, depending on whether the change is evaluated as positive or negative (Colquitt et al., 2013). Based on this, an automated decision process can be expected to be perceived differently than a non automated decision process especially when it comes to procedural fairness (Hypothesis 1) and trust (Hypothesis 2). However, since automation does not affect who makes the decision, the expectation regarding the bonus payment should remain the same, and therefore the expectation of what is being decided should also remain unchanged (Hypothesis 3).

**Hypothesis 1** *Under a non automated decision process, agents perceive procedural fairness differently than under an automated decision process.*

**Hypothesis 2** *Under a non automated decision process, agents perceive a different level of trust than under an automated decision process.*

**Hypothesis 3** *Under a non automated decision process, agents have the same expectation about the performance threshold to receive a bonus than under an automated decision process.*

Research has not only shown that performance is influenced by situational factors such as fairness and trust (e.g., Fehr et al., 2007), but also control (e.g., Fehr and Rockenbach, 2003; Fehr and List, 2004; Falk and Kosfeld, 2006; Fehr and Schmidt, 2007; Kajackaite and Werner, 2015). These aspects in particular can be influenced by automation. Automation is based on quantifiable parameters, but different parameters can be quantified more or less well. Hence, automation can lead to decontextualization if, for example, specific parameters, which are difficult to quantify, are no longer included. If automation is perceived as decontextualizing and quantifying the process, people might react to it by changing their behavior, depending on whether they perceive it as a positive or negative change. Hence, the agent's performance is expected to differ depending on whether a non automated or automated decision process is used (Hypothesis 4).

**Hypothesis 4** *Under a non automated decision process, agents exert a different level of performance than under an automated decision process.*

Furthermore, agents might also value the fact that the principal is directly controlling whether the bonus is paid or not and feel more appreciated if the decision is made knowing the actual performance. This enables direct adjustments, a factor found to influence procedural fairness

(e.g., Colquitt, 2001). Research on intention-based reciprocity by Rabin (1993), Dufwenberg and Kirchsteiger (2004), and Falk and Fischbacher (2006), also shows that people tend to reward kind intentions and punish unkind ones. By actively choosing an automated decision process, the principal abstains from directly controlling whether a bonus is going to be paid, as the principal decides not to know the agent's performance before determining the performance threshold to receive a bonus. The principal thus actively shows an information avoidance behavior. In this respect, the principal's decision to use an automated decision process could be perceived as a lack of interest in or appreciation for the agent's performance. Therefore, the agents might reciprocate by choosing a lower performance level if the principal decides to use an automated decision process. However, if an automated third party instead of the principal decides whether to use an automated decision process or a non automated decision process, the agent's reaction based on reciprocity can be expected to be less pronounced because of the lack of a direct counterpart to be held accountable for the decision.

Based on the considerations above, agents are expected to show a higher performance level if the principal decides to use a non automated decision process compared to when an automated third party chooses the non automated decision process. Correspondingly, agent's performance is expected to be lower if the principal decides to use an automated decision process compared to when an automated third party chooses the automated decision process (Hypothesis 5).

**Hypothesis 5** *If the principal decides on the process, agents show a*

*(i) higher performance in a non automated decision process, and*

*(ii) lower performance in an automated decision process*

*than if a third party decides on the process.*

# 5. Results

Overall, 508 participants (43.3% female) contributed to the study.[10] As Table 2 shows, in treatment *HUMAN*, 256 participants (41.7% female) and in treatment *SYSTEM*, 252 participants (54.7% female) took part in the experiment. Half of the participants took the role of an agent, and the other half took the role of a principal (254 participants each). In treatment *HUMAN*, the principals chose the non automated process more often than the automated process, which can be seen as an indication for algorithm aversion.

[10] Three participants dropped out during the experiment. These are not included in the numbers anymore. The participants that dropped out during the experiment were moved by the experimenter during the experiment so that their matched participants were able to finish the experiment. For the results, these participants and their matches were removed as the behavior of the experimenter might have influenced the decisions of the matched participants. The questions (e.g., on fairness and trust) were answered by only 240 of the 254 agents, as 14 agents left the experiment directly before answering the questions in Stage 4. In the treatment *HUMAN*, a total of 9 agents did not answer the questions (3 in automated, 6 in non automated). In the treatment *SYSTEM*, a total of 5 agents did not answer the questions (3 in automated, 2 in non automated). This is a well-known challenge for online experiments, as control over participants is limited.

The participants were on average 37 years old. The study took about 10 minutes to complete, and the participants earned on average $1.73. As a between-subjects design was used for the treatments and a within-subject design to elicit the agents performance, the data for statistical tests is independent between but dependent within treatments.

In the following Section, the agents' perceived fairness of and trust in the process, as well as their expectation about the performance threshold, is analyzed followed by a comparison of the performance in the non automated decision process and the automated decision process within and between treatments.[11]

| | non automated | automated | total |
|---|---|---|---|
| *HUMAN* | 200 | 56 | 256 |
| *SYSTEM* | 124 | 128 | 252 |
| total | 324 | 184 | 508 |

There was an equal distribution between principals and agents for each of the numbers presented above.

Table 2: Number of participants per process type and treatment.

## 5.1. Hypothesis 1: Perceived fairness of the process

As Table 3 shows, the fairness assessments by agents do not differ significantly between a non automated decision process and an automated decision process. Thus, agents do not perceive one process to be fairer than the other, and Hypothesis 3, i.e., that agents will perceive procedural fairness differently in both processes, cannot be confirmed.[12]

| | non automated | automated |
|---|---|---|
| | ($p$ =0.7725) | |
| YES | 83.77 (129) | 81.40 (70) |
| NO | 16.23 (25) | 18.60 (16) |

The Table shows the percentage of agents who perceived the process to be (not) fair and $p$-values for a Chi-squared test. The absolute number is shown in parenthesis.

Table 3: Agents' answer to the question: "Do you consider the procedure to get the bonus to be fair?" (see Question 6 from Appendix A.8).

## 5.2. Hypothesis 2: Perceived trust in the process

Table 4 shows that most agents expected to get a bonus, but agents' expectations about receiving a bonus did not differ significantly between a non automated decision process and

[11] An analysis of the principals' behavior can be found in Section 5.3 and Appendix A.5.

[12] The fairness assessment by the principal can be found in Table 14 in Appendix A.6. The fairness assessment by agents and principals by treatment can be found in Table 9 in Appendix A.1.

an automated decision process.[13] Hence, Hypothesis 2, i.e., that agents depending on the process perceive a different level of trust, cannot be confirmed.

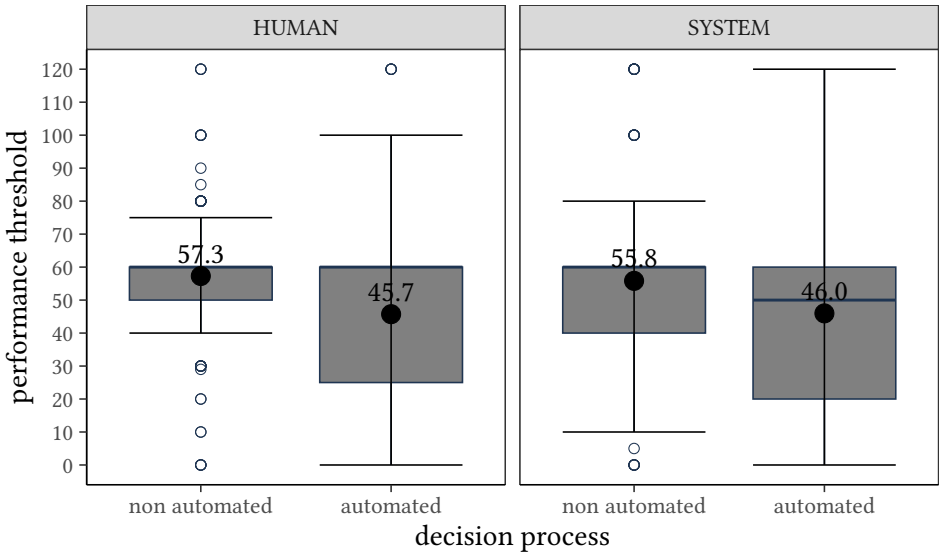
| | non automated | automated |
|---|---|---|
| | ($p$ =0.6398) | |
| Strongly agree | 14.3 (22) | 11.6 (10) |
| Agree | 71.4 (110) | 70.9 (61) |
| Disagree | 12.3 (19) | 14 (12) |
| Strongly disagree | 1.9 (3) | 3.5 (3) |

The Table shows the percentages of agents' perceptions and $p$-value for a Cochran-Armitage test for trend. The absolute number is shown in parenthesis.

Table 4: Agents' agreement to the statement: "I think that I will get a bonus" (see Question 5 from Appendix A.8).

## 5.3. Hypothesis 3: Expectation about the performance threshold

As Figure 1 shows and Table 5 confirms, agents in both treatments expect the performance threshold required to receive to be significantly higher in the non automated decision process than in the automated decision process. Hence, Hypothesis 3, i.e., that agents will have the same expectation about the performance threshold in both processes, cannot be confirmed.




Filled dots represent means; lines represent medians.

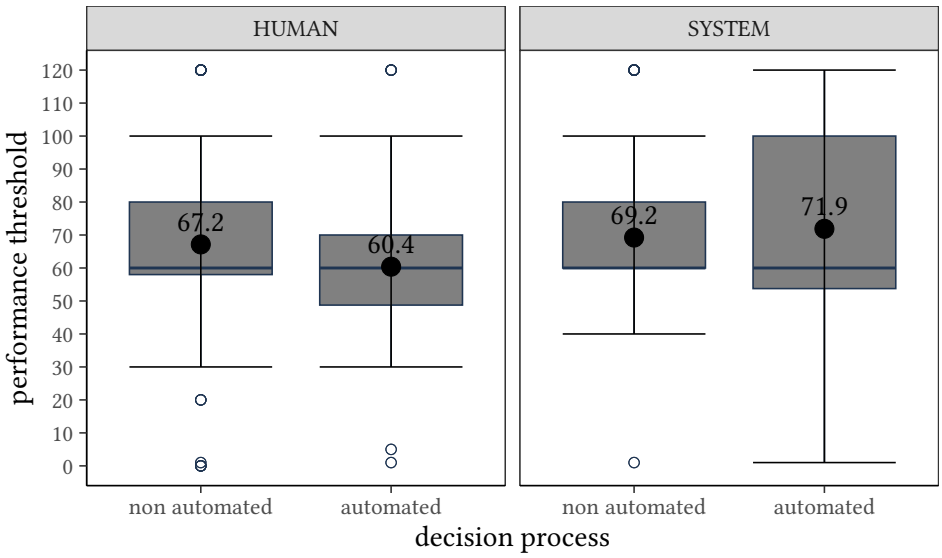
Figure 1: Box-and-whisker plots for agents' expectation about the performance threshold set by the principal (see Questions 3 and 4 from Appendix A.8).

[13]The perceived trust to receive a bonus by treatment can be found in Table 10 in Appendix A.2.

| | *HUMAN* | *SYSTEM* |
|---|---|---|
| non automated - automated | $\Delta = 11.57$ ($p$ =0.0000) | $\Delta = 9.86$ ($p$ =0.0001) |
| | $\Delta = 10.71$ ($p$ =0.0000) | |

The Table shows differences between the performance threshold expectations ($\Delta = \ldots$) and $p$-values for a two-sided Wilcoxon signed-rank test/Wilcoxon rank-sum test if the difference could be zero.

Table 5: Differences in the agent's expectations about the performance threshold (see Questions 3 and 4 from Appendix A.8).




Filled dots represent means, lines represent medians.

Figure 2: Box-and-whisker plots for performance thresholds set by principals.

Interestingly, agents anticipate a different performance threshold in both processes, but the actual threshold by the principals does not differ significantly. Figure 2 shows and Table 6 confirms, the performance threshold set by the principals in the automated decision process does not differ significantly from the performance threshold set in the non automated decision process in both treatments.

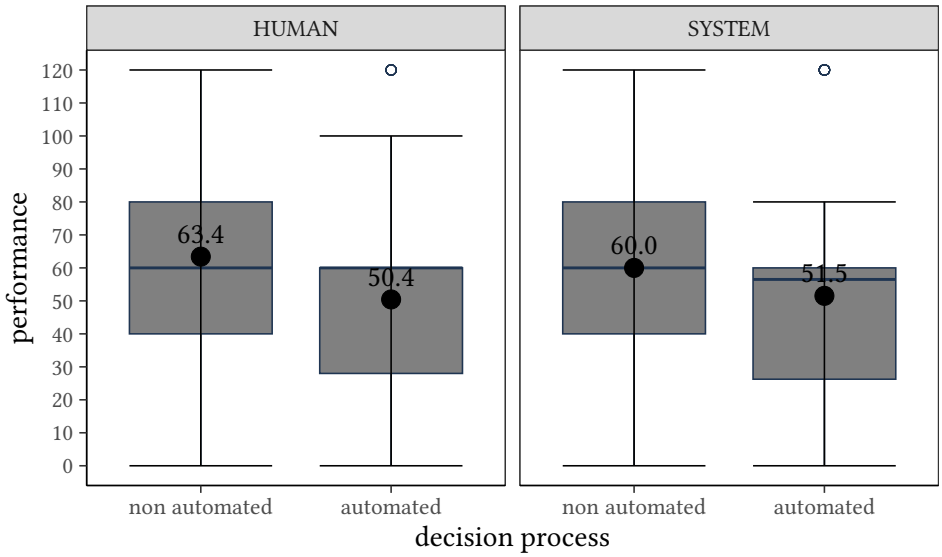
| | *HUMAN* | *SYSTEM* |
|---|---|---|
| non automated - automated | $\Delta$ = 6.76 | $\Delta$ = -2.65 |
| | ($p$ =0.1953) | ($p$ =0.8511) |
| non automated | $\Delta$ = -2.08 | |
| | ($p$ =0.9565) | |
| automated | $\Delta$ = -11.49 | |
| | ($p$ =0.2283) | |

The Table shows the performance threshold difference ($\Delta = \ldots$) and $p$-values for a two-sided Wilcoxon signed-rank test/Wilcoxon rank-sum test if the difference could be zero.

Table 6: Differences in the principals' performance threshold set between the non automated decision process and automated decision process.

The performance thresholds set in automated decision process in treatment *SYSTEM*, however, are more dispersed than in the other conditions. Principals in treatment *HUMAN*, deciding on their own which approach to use, also do not set a significantly different performance threshold than principals in treatment *SYSTEM*, where an automated third party randomly decides which approach to use.

## 5.4. Hypothesis 4: Performance depending on automation




Filled dots represent means, lines represent medians.

Figure 3: Box-and-whisker plots for agents' performance.

| | *HUMAN* | *SYSTEM* |
|---|---|---|
| non automated - automated | $\Delta = 12.98$ ($p$ =0.0000) | $\Delta = 8.47$ ($p$ =0.0001) |
| | $\Delta = 10.74$ ($p$ =0.0005) | |

The Table shows the performance difference ($\Delta = \ldots$) and and $p$-values for a two-sided Wilcoxon signed-rank test/Wilcoxon rank-sum test. The tests report no problem regarding the frequency of ties.

Table 7: Differences in the agents' performance between treatments.

As Figure 3 shows, agents' performance in the non automated decision process is higher than in the automated decision process in both treatments.[14] Table 7 shows the performance difference between the non automated decision process and the automated decision process, and confirms that the agents' performance is significantly higher in the non automated decision process than in the automated decision process in both treatments. Hence, Hypothesis 4, i.e., that the performance level differs between a non automated and an automated process, can be confirmed.

## 5.5. Hypothesis 5: Performance depending on who decides to automate

According to Hypothesis 5.(i), the agents' performance should be higher in the non automated decision process and, according to Hypothesis 5.(ii), lower in the automated decision process in treatment *HUMAN* than in treatment *SYSTEM*. Indeed, this is what we see in Figure 3. Agents' performance in treatment *HUMAN* in the non automated (automated) decision is higher (lower) than in treatment *SYSTEM*. However, as Table 8 shows, the difference is not statistically significant either in the non automated or the automated decision. Thus, Hypothesis 5.(i) and Hypothesis 5.(ii) cannot be confirmed.

| | *HUMAN - SYSTEM* |
|---|---|
| non automated | $\Delta = 3.42$ ($p = 0.1194$) |
| automated | $\Delta =$ -1.09 ($p$ =0.5028) |

The Table shows the performance difference ($\Delta = \ldots$) and $p$-values for a one-sided Wilcoxon rank-sum test. The tests report no problem regarding the frequency of ties.

Table 8: Performance differences between treatments.

[14] A detailed analysis of the agents' performance can be found in Table 12 in Appendix A.4.

# 6. Discussion

The experiment has two main limitations that need to be addressed. First, the experiment was designed as an online experiment. As all interaction took place via a computer interface, the difference between a human decision and a decision made by an automated third party is quite subtle. Thus, the divergence between processes might not have been salient enough to the participants, explaining the similarity of performance between the two processs for half of the participants. One can also assume that the number of participants performing the same in both processes may be lower in a laboratory experiment as factors such as quantification, reductionism and rigidity become more salient to the participants. In addition, the psychological distance between the principal and the agent in an online experiment could also influence procedural justice perceptions when it comes to automation. Second, the experiment models a one-shot interaction. A one-shot interaction might not apply when participants interact over a longer period because participants might become more aware of dissimilarities between the automated and non human processes. In interactions over a longer period or a sequence of decisions, possible differences between decisions due to a learning and adaption effect might occur. It might be that when people interact over a longer time, they evaluate certain characteristics such as high consistency and lack of bias but no possibility to voice opinions or concerns about the automated process differently. Many aspects of how automation influences performance remain unknown. Further research should look at repeated interaction situations and how proximity between principal and agent affects performance when management processes are automated. It would also be interesting to shed light on the effects of subjective or objective control over the bonus payment.

# 7. Conclusion

The adoption of algorithmic automation as a tool for managerial decision-making is increasingly expanding, covering a wide range of activities including candidate selection, optimization of incentive compensation schemes, performance evaluation, and the execution of performance appraisal decisions. Thereby, it is crucial to bear in mind that the progressive automation leads to a transformation of the automated processes themselves. Despite the fact that process automation does not alter who makes the decisions, it transforms how a decision is made. Automation simplifies and streamlines the decision-making process. As a result, automated management processes rely less on an individual's judgment and more on pre-established rules, reducing the level of personalization and introducing a higher degree of rigidity. As management tasks become more automated, it is therefore important to understand how people perceive the increased automation, and how they react to it.

This paper analyzes how process automation impacts perceived process fairness, trust expectations, and human performance. Using an online experiment, I examine the consequences of automating a bonus payment process within a modified principal-agent game setting. Analogous to the concept of process automation, where decision parameters are predetermined to configure an algorithm, the bonus payment expectations were established either with or without prior knowledge of the actual performance (ex post or ex ante), thus

modeling the inherent inflexibility of an automated system. In addition, I investigate the extent to which it matters whether it is the immediate supervisor or a third party who makes the decision to automate,

The experiment provides two sets of results. First, automated processes can lead to a decline in performance. The results show a significant effect of automated decisions on performance. More specifically, I found that performance is significantly lower under an automated than under a non automated decision process. Furthermore, performance is higher in the non automated decision process and lower in the automated decision process if a human rather than if an automated third party decides to use an automated process. Although this difference is not statistically significant, it is in the expected direction and can serve as a starting point for examining the extent to which the decision of who automates affects performance. In this vein, another noteworthy additional aspect is that participants exhibited a higher likelihood of expecting to receive a bonus if a human rather than if an automated third party decides to use an automated process. Second, contrary to expectations, it is not fairness and trust issues that are decisive, but rather divergent expectations. More precisely, while the perceived fairness of both processes as well as the trust to receive a bonus did not differ significantly between both processes, the expectation about the minimum performance level required to receive a bonus did. Agents expected the performance expectation to be significantly higher in the non automated than in the automated process independent of who decided to use the automated process. However, the principals expectations in a non automated decision process and in an automated decision process are more or less the same. In other words, the agents expected the performance expectation to receive a bonus to be higher in a process determined after the performance was delivered (ex post) compared to a process determined beforehand (ex ante). Given that the decision-maker remains unchanged, but only the decision-making process undergoes modification, one would expect that the expectations would remain unchanged. This is however not the case as the results reveal a significant disparity in expectations between the two processes. Put simply, the result indicate that there are divergent expectations among individuals in an automated process when compared to a non-automated process, and the subsequent change in willingness to perform can be attributed to these altered expectations. This finding is noteworthy as it demonstrates that divergent expectations concerning the necessary performance have a discernible impact on the diminished performance observed in an automated process. Although performance was not significantly influenced by whether a person or a random mechanism decided to employ a specific process, the results suggest that the decision to automate a process, whether undertaken by the immediate supervisor or by the organization as an independent entity, does impact performance. Moreover, the findings also reveal signs of algorithm aversion, as the principals exhibited a higher preference for the non automated process compared to the automated process. For future research it would be therefore intriguing to investigate whether this algorithm aversion undergoes any changes in the future as process automation becomes increasingly prevalent in our daily lives.

The contribution of this paper goes beyond a mere examination of the unintended consequences arising from automation in management tasks as the main finding of this paper is that automation can have significant downstream effects on individual performance that is driven by an unintended modification of one of the underlying factors linked to the process.

The primary breakthrough therefore lies in my demonstration that automation has the capacity to modify expectation formation within automated processes, leading to a substantial downstream effect on performance. This finding does not only contribute to the evolving research field on the unintended consequences of automation but also strengthens the existing economic literature on incomplete contracts as a more or less rigid process can essentially be regarded as a more or less rigid contract. By investigating the responses to the heightened rigidity of personal decisions resulting from automation, the paper also expand the scope of research on procedural justice. The paper contributes to the existing literature by highlighting the significance of human involvement, which is perceived as enhancing the utilization of social information and situational context when it comes to fairness issues.

Given that I can pinpoint the reason for the decrease in performance – divergent performance expectations – the following lessons can be learned. In order to mitigate unintended consequences of automation, the adoption of automation within organizations should be evaluated not only based on cost savings but also considering its social implications. In order to mitigate potential hidden costs, such as a decline in performance, that may arise from automation, it is essential to recognize that automation not only mechanizes a process but also influences and modifies the process itself. Consequently, placing a high priority on comprehensive communication becomes crucial, as it serves as a mechanism to mitigate the potential development of misguided expectations. In light of this, organizations pursuing efficiency gains through the automation of social management tasks should be mindful of the potential negative impact on employee performance arising from divergent expectations during the automation process.

# References


Aharoni, E. and Fridlund, A. J. (2007). Social reactions toward people vs. computers: How mere lables shape interactions. *Computers in Human Behavior*, 23(5):2175–2189.

Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

Bai, B., Dai, H., Zhang, D. J., Zhang, F., and Hu, H. (2022). The impacts of algorithmic work assignment on fairness perceptions and productivity: Evidence from field experiments. *Manufacturing & Service Operations Management*, 24(6):3060–3078.

Bauer, P. C. and Freitag, M. (2018). Measuring trust. *The Oxford handbook of social and political trust*, 15.

Beauvois, J.-L. and Joule, R. (1996). *A radical dissonance theory*. Taylor & Francis, London, GB.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

Berger-Böcker, S. (2019). Survey analysis: Artificial intelligence is on the rise for HR. Retrieved February 23, 2023, from https://www.gartner.com/en/documents/3980246.

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

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193.

Brynjolfsson, E. and Mcafee, A. (2017). Artificial intelligence, for real. *Harvard business review*, 1:1–31.

Burdin, G., Halliday, S., and Landini, F. (2018). The hidden benefits of abstaining from control. *Journal of Economic Behavior & Organization*, 147:1–12.

Castelo, N. (2019). *Blurring the line between human and machine: marketing artificial intelligence*. Columbia University.

Charness, G., Gneezy, U., and Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, 149(3):74–87.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree — An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

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

Chugunova, M. and Sele, D. (2022). An interdisciplinary review of the experimental evidence on how humans interact with machines. *Journal of Behavioral and Experimental Economics*, page 101897.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, 86(3):386.

Colquitt, J. A., Scott, B. A., Rodell, J. B., Long, D. M., Zapata, C. P., Conlon, D. E., and Wesson, M. J. (2013). Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *Journal of Applied Psychology*, 98(2):199.

Cormier, D., Newman, G., Nakane, M., Young, J. E., and Durocher, S. (2013). Would you do as a robot commands? An obedience study for human-robot interaction. In *International Conference on Human-Agent Interaction*, pages 1–3.

Danilov, A. and Chugunova, M. (2022). Use of digital technologies for HR management in germany: Survey evidence. *MPRA Paper*, (111530).

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1):114.

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. *Management Science*, 64(3):1155–1170.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2):268–298.

Ellingsen, T. and Johannesson, M. (2005). Trust as an incentive. *Stockholm School of Economics Working Paper mimeo*.

Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2):293–315.

Falk, A. and Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, 96(5):1611–1630.

Fehr, E., Klein, A., and Schmidt, K. M. (2007). Fairness and contract design. *Econometrica*, 75(1):121–154.

Fehr, E. and List, J. A. (2004). The hidden costs and returns of incentives — trust and trustworthiness among ceos. *Journal of the European Economic Association*, 2(5):743–771.

Fehr, E. and Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature*, 422(6928):137–140.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fehr, E. and Schmidt, K. M. (2007). Adding a stick to the carrot? The interaction of bonuses and fines. *American Economic Review*, 97(2):177–181.

Fisher, A. (2019). An algorithm may decide your next pay raise. Retrieved February 27, 2023, from `https://fortune.com/2019/07/14/artificial-intelligence-workplace-ibm-annual-review/`.

Greenfield, R. (2018). Your raise is now based on next year's performance. Retrieved February 27, 2023, from `https://financialpost.com/executive/your-raise-is-now-based-on-next-years-performance`.

Hertz, N. and Wiese, E. (2019). Good advice is beyond all price, but what if it comes from a machine? *Journal of Experimental Psychology: Applied*, 25(3):386.

Höddinghaus, M., Sondern, D., and Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, 116:106635.

Hooper, R. S., Galvin, T. P., Kilmer, R. A., and Liebowitz, J. (1998). Use of an expert system in a personnel selection process. *Expert Systems with Applications*, 14(4):425–432.

Huang, M.-H. and Rust, R. T. (2018). Artificial intelligence in service. *Journal of service research*, 21(2):155–172.

Jussupow, E., Benbasat, I., and Heinzl, A. (2020). Why are we averse towards algorithms? a comprehensive literature review on algorithm aversion. *Research Papers*, 168.

Kaibel, C., Koch-Bayram, I., Biemann, T., and Mühlenbock, M. (2019). Applicant perceptions of hiring algorithms-uniqueness and discrimination experiences as moderators. In *Academy of Management Proceedings*, volume 2019, page 18172. Academy of Management Briarcliff Manor, NY 10510.

Kajackaite, A. and Werner, P. (2015). The incentive effects of performance requirements – a real effort experiment. *Journal of Economic Psychology*, 49:84–94.

Kampkötter, P. (2017). Performance appraisals and job satisfaction. *The International Journal of Human Resource Management*, 28(5):750–774.

Kaur, N. and Sood, S. K. (2017). A game theoretic approach for an iot-based automated employee performance evaluation. *IEEE Systems Journal*, 11(3):1385–1394.

Konow, J. (1996). A positive theory of economic fairness. *Journal of Economic Behavior & Organization*, 31(1):13–35.

Kusterer, D. and Sliwka, D. (2022). Social preferences and rating biases in subjective performance evaluations. *IZA Discussion Paper*, (No. 15496).

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. *Computers in Human Behavior*, 123:106878.

Lawler, J. J. and Elliot, R. (1996). Artificial intelligence in HRM: An experimental study of an expert system. *Journal of Management*, 22(1):85–111.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1):2053951718756684.

Levanthal, G. S. (1980). What should be done with equity theory. In *Social exchange: Advances in theory and research*, pages 27–55. Plenum press New York.

Levy, P. E. and Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, 30(6):881–905.

Liem, C., Langer, M., Demetriou, A., Hiemstra, A. M., Sukma Wicaksana, A., Born, M. P., and König, C. J. (2018). Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. In *Explainable and Interpretable Models in Computer Vision and Machine Learning*, pages 197–253. Springer.

Markets and Markets (2020). Business process management market by component, deployment type, organization size, business function (sales and marketing, HRM, Procurement and SCM, and Customer Service Support), industry, and region - Global Forecast to 2025. Retrieved February 27, 2023, from `https://www.marketsandmarkets.com/Market-Reports/business-process-management-market-157890056.html`.

Meijerink, J., Boons, M., Keegan, A., and Marler, J. (2021). Algorithmic human resource management: Synthesizing developments and cross-disciplinary insights on digital HRM. *The International Journal of Human Resource Management*, 32(12):2545–2562.

Nagtegaal, R. (2021). The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly*, 38(1):101536.

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. *Organizational Behavior and Human Decision Processes*, 160:149–167.

Ötting, S. K. and Maier, G. W. (2018). The importance of procedural justice in human–machine interactions: Intelligent systems as new decision agents in organizations. *Computers in Human Behavior*, 89:27–39.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, 83(5):1281–1302.

Rabin, M. (1995). Moral preferences, moral constraints, and self-serving biases. *Department of Economics UCB Working Paper No. 95-241.*

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., et al. (2019). Machine behaviour. *Nature*, 568(7753):477–486.

Raisch, S. and Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of management review*, 46(1):192–210.

Riberolles, H. (2020). Modernisation of incentive compensation: from simple commission rates to the most sophisticated calculation algorithms. Retrieved February 28, 2023, from `https://www.primeum.com/en/blog/modernisation-of-incentive-compensation-from-simple-commission-rates-to-the-most-sophisticated-calculation-algorithms`.

Schmelz, K. and Ziegelmeyer, A. (2015). Social distance and control aversion: Evidence from the internet and the laboratory. *Thurgau Institute of Economics and Department of Economics at the University of Konstanz Research Paper Series, No.100.*

Stone, H. S. (1971). *Introduction to computer organization and data structures.* McGraw-Hill, New York, NY.

Sturgis, P. and Smith, P. (2010). Assessing the validity of generalized trust questions: What kind of trust are we measuring? *International Journal of Public Opinion Research*, 22(1):74–92.

Thibaut, J. W. and Walker, L. (1975). *Procedural justice: A psychological analysis.* L. Erlbaum Associates.

Upadhyay, A. K. and Khandelwal, K. (2018). Applying artificial intelligence: implications for recruitment. *Strategic HR Review*, 17(5):255–258.

Van Esch, P., Black, J. S., and Ferolie, J. (2019). Marketing ai recruitment: The next phase in job application and selection. *Computers in Human Behavior*, 90:215–222.

Villeval, M. C. (2020). Performance feedback and peer effects. *GLO Discussion Paper*, (No. 482).

Votto, A. M., Valecha, R., Najafirad, P., and Rao, H. R. (2021). Artificial intelligence in tactical human resource management: A systematic literature review. *International Journal of Information Management Data Insights*, 1(2):100047.

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., and Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: A systematic review. *The International Journal of Human Resource Management*, 33(6):1237–1266.

Wang, R., Harper, F. M., and 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*, pages 1–14.

Waytz, A. and Norton, M. I. (2014). Botsourcing and outsourcing: Robot, british, chinese, and german workers are for thinking—not feeling—jobs. *Emotion*, 14(2):434.

Yen, I.-L., Bastani, F., Huang, Y., Zhang, Y., and Yao, X. (2017). Saas for automated job performance appraisals using service technologies and big data analytics. In *2017 IEEE International Conference on Web Services (ICWS)*, pages 412–419. IEEE.

## Statements and Declaration

This work was funded by the Max Planck Society through the International Max Planck Research School on Adapting Behavior in a Fundamentally Uncertain World. The author has no relevant financial or non financial interests to disclose.

# A. Appendix

This Section contains additional information on the interfaces and questions used in the treatments. I also present further analyses of the data I collected in addition to the data used to test my hypotheses. Data and methods are available upon request.

## A.1. Fairness Assessment

| | | *HUMAN* | | *SYSTEM* | |
|---|---|---|---|---|---|
| | | non automated | automated | non automated | automated |
| Agent | YES | 84.04 | 88.00 | 83.33 | 78.69 |
| | NO | 15.96 | 12.00 | 16.67 | 21.31 |
| Principal | YES | 85.00 | 78.57 | 87.10 | 79.69 |
| | NO | 15.00 | 21.43 | 12.90 | 20.31 |

The Table shows the percentage of agents and principals who perceived the process to be (not) fair.

Table 9: Agents' and principals' answer to the following question: "Do you consider the procedure to get the bonus to be fair?" (see Question 6 from Appendix A.8).

## A.2. Trust Assessment

| | HUMAN | SYSTEM |
|---|---|---|
| Strongly agree | 9.20 | 17.40 |
| Agree | 77.30 | 65.30 |
| Disagree | 10.10 | 15.70 |
| Strongly disagree | 3.40 | 1.70 |

The Table shows the percentage of agents perceptions.

Table 10: Agents' agreement [%] to the following statement: "I think that I will get a bonus" (see Question 5 from Appendix A.8).

## A.3. Performance occurrence according to social preferences

Table 11 shows four social preferences, indicating that participants might prefer certain productive activities of the agent over others. Assuming purely selfish preferences, the principal would not care whether the agent receives a bonus or not. Under this condition, agents maximize their payoff by choosing a transfer of $x = 0$. Assuming efficiency preferences, the agent and the principal strive to maximize overall social welfare. The agent would transfer $x = 120$ points, and the principal ensures a bonus payment of $b^* = 120$. Assuming that the agent expects the probability for receiving a bonus to depend on the transfer, a more detailed analysis is required. If the principal attaches importance to a fair split, the principal might demand

| Social preference | Performance preference ($x$) |
|---|---|
| Selfishness | 0 |
| Efficiency | 120 |
| Fair split | $\sim 60$ |
| Equality | $\sim 80$ |

The Table shows possible social preferences and the corresponding preferred performance.

Table 11: Preferred performance($x$) according to social preferences

an equal split of the agent's initial endowment and therefore prefers the agent to choose a productive activity of $x \approx 60$. If the principal's utility is negatively influenced by an unequal outcome (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), the principal would want the agent to choose a productive activity roughly around $x \approx 80$, ensuring an overall equal outcome. An utility-maximizing agent would therefore anticipate the preferences of the principal and choose a productive activitiy that matches the anticipated expectation set by the principal.[15]

## A.4. Relative frequency of the agents' transfer decision

| | HUMAN | SYSTEM |
|---|---|---|
| less | 42.20 | 30.20 |
| more | 7.00 | 7.10 |
| same | 50.80 | 62.70 |

The Table shows the percentage of agents transferring the same, more, or less in an automated decision process than in a non automated decision process by treatment.

Table 12: Agents' transfer decisions [%].

Table 12 shows the relative frequency of agents who transferred less, more, or the same in an automated decision process than in a non automated decision process. The Table reveals that around half of the agents transferred the same in a non automated decision process as in an automated decision process in both treatments. Nevertheless, around 40% of the agents transferred fewer points in an automated decision process than in a non automated decision process in treatment *HUMAN*, and slightly less than one-third of the agents did so in treatment *SYSTEM*.

[15] Models of social image concerns (e.g., Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009) and concepts of self-perception maintenance (e.g., Rabin, 1995; Beauvois and Joule, 1996) suggest that individuals perceive an unpleasant tension or disutility if their actions cause harm to their social concept and/or self-concept of being a kind and fair individual. Thus, the productive activity of the agent may also be infuenced by self- and social-image concerns, as well as risk preferences.

## A.5. Analysis of the principals' behavior

As Table 13 shows, the vast majority of the principals decided to use a non automated instead of an automated decision in treatment *HUMAN*, where the principals were able to choose.

| | HUMAN | SYSTEM |
|---|---|---|
| non automated | 78.10 | 49.20 |
| automated | 21.90 | 50.80 |

The Table shows the percentage of principals choosing a non automated decision process or automated decision process.

Table 13: Process choices by principals and the system[%].

## A.6. Participants' propensity for procedural fairness, risk and trust

As Table 14 shows, the fairness assessments by principals do not differ significantly between a non automated decision process and an automated decision process. Thus, principals do not perceive one process to be fairer than the other.

| | non automated | automated |
|---|---|---|
| | ($p$ =0.2480) | |
| YES | 85.80 | 79.35 |
| NO | 14.20 | 20.65 |

The Table shows the percentage of principals who perceived the process to be (not) fair and $p$-values for a chi-square test.

Table 14: Principals' answer to the question: "Do you consider the procedure to get the bonus to be fair?" (see Question 6 from Appendix A.8).

All participants were asked if they are a person who is willing to take risks or tries to avoid taking risks (see Question 7) and if they would say that most people can be trusted or that you cannot be too careful in dealing with other people (see Question 8). Willingness to take risks was measured by a continuous scale from "not at all willing to take risks" (0) to "very willing to take risks" (10). The level of trust was measured by a continuous scale from "can't be too careful" (0) to "most people can be trusted" (10). As Table 15 shows, agents and principals were slightly risk- averse and somewhat concerned about the trustworthiness of other people.

| | Agent | Principal |
|---|---|---|
| Risk | ∅ = 4.36 | ∅ = 4.53 |
| | (2.59) | (2.3) |
| Trust | ∅ = 4.91 | ∅ = 5.04 |
| | (2.55) | (2.41) |

The Table shows the means for risk and trust (∅ = . . .) and the corresponding standard deviations (in brackets).

Table 15: Means and standard deviations for levels of risk and trust.

## A.7. Instructions

*Note: The following instructions are for treatment HUMAN. Differences for treatment SYSTEM are added in italic.*

This HIT [Human Intelligence Task] is an economic experiment. Please read the following instructions carefully. The instructions provide you with all the information required for participating in the experiment. You will receive $0.50 USD for participating in the experiment (paid only if you finish the experiment). Your final payoff is the $0.50 USD for participating in the experiment plus the amount earned during the experiment. You will earn at least the $0.50 USD for participating in the experiment. In the experiment, the currency used is points. Your points will be converted to USD at the end of the experiment using a conversion rate of **1 point = $0.01 USD**.

**General setup**

In this experiment, you are matched with another human participant. You will play in a group of two. All decisions are made anonymously. No participant knows with whom (s)he is matched. During the experiment, the members of the group are called "participant A" and "participant B". The roles are randomly assigned.

**The experiment**

Participant A starts with 120 points at the beginning of the experiment. Participant B starts with no points. Each participant has to make a decision during the experiment. The decisions are explained below. Please read the explanations for both participants as both decisions will affect the number of points you will earn.

**Participant A's decision:**

Participant A has to decide how many points (s)he wants to transfer to participant B. The points transferred to participant B are doubled by the experimenter, meaning each point transferred to participant B reduces the points of participant A by one point but increases the points of participant B by two points.

**Participant B's decision:**

Before participant B knows what participant A transferred, participant B *[the system]* selects an approach. The two possible approaches (Approach BLUE or Approach GREEN) are explained below.

In each approach participant B has to decide if participant A should be given a **bonus of 120 points** for a **"very good transfer"**. The bonus is paid by the experimenter and does not reduce the points of participant B. The difference between the two approaches is how participant B determines the minimum amount (threshold) that participant A has to transfer to get a bonus.

In Approach BLUE, participant B **knows** the amount transferred by participant A when determining the threshold. The decision screen will look like this:

| |
|---|
| You *[the system]* decided to use Approach BLUE.<br><br>**Participant A has transferred X of 120 points to you.**<br><br>If participant A has transferred at least the threshold amount (s)he gets a bonus of 120 points (paid by the experimenter).<br>Please indicate your threshold here:<br><br>……Points |

In Approach GREEN, participant B **DOES NOT know** the amount transferred by participant A when determining the threshold. The decision screen will look like this:

| |
|---|
| You *[the system]* decided to use Approach GREEN.<br><br>If participant A has transferred at least the threshold amount (s)he gets a bonus of 120 points (paid by the experimenter).<br>Please indicate your threshold here:<br><br>……Points |

**Further note:**
Participant A has different fields to enter amounts in case Approach BLUE or Approach GREEN is used.
***Some examples:***

- ***Example 1:*** Participant A transfers 0 points to participant B. Participant A will have 120 points (120 - 0) plus eventually a bonus of 120 points. Participant B will have 0 points (0 x 2). In addition, both participants receive $0.50 USD for participating.
- ***Example 2:*** Participant A transfers 40 points to participant B. Participant A will have 80 points (120 - 40) plus eventually a bonus of 120 points. Participant B will have 80 points (40 x 2). In addition, both participants receive $0.50 USD for participating.

- ***Example 3:*** Participant A transfers 80 points to participant B. Participant A will have 40 points (120 - 80) plus eventually a bonus of 120 points. Participant B will have 160 points (80 x 2). In addition, both participants receive $0.50 USD for participating.
- ***Example 4:*** Participant A transfers 120 points to participant B. Participant A will have 0 points (120 - 120) plus eventually a bonus of 120 points. Participant B will have 240 points (120 x 2). In addition, both participants receive $0.50 USD for participating.

Before clicking "Next" please make sure you have read and understood the instructions. After clicking "Next" we will match you with the next person starting the experiment. This might take some time.

## A.8. Questions

*Note: All participants were asked to complete a questionnaire. The questions were asked right after the decision and before the final outcome was announced. The answer method used is presented in brackets. Apart from the first five questions, which were only presented to agents, all questions were asked to agents and principals.*

1. Why did you choose to transfer the amount you have chosen to participant B in Approach BLUE (participant B **knows** how much you transferred)? [Open Question] *(For the answers given, see online data set.)*
2. Why did you choose to transfer the amount you have chosen to participant B in Approach GREEN (participant B **does not know** how much you transferred)? [Open Question] *(For the answers given, see online data set.)*
3. What do you think is the minimum amount you would have had to transfer to get the bonus if participant B decided to use Approach BLUE (participant B **knows** how much you transferred)? [Integer from 0 to 120 points] *(For an analysis of the answers given, see Section 5.4.)*
4. What do you think is the minimum amount you would have had to transfer to get the bonus if participant B decided to use Approach GREEN (participant B **does not know** how much you transferred)? [Integer from 0 to 120 points] *(For an analysis of the answers given, see Section 5.4.)*
5. How much do you agree with this statement: 'I think that I will get a bonus.'? *(For an analysis of the answers given, see Appendix A.2.)*
6. Do you consider the procedure to get the bonus to be fair? *(For an analysis of the answers given, see Appendix A.1.)*
7. How do you see yourself: Are you a person who is willing to take risks or do you try to avoid taking risks? Please select a number on a scale from 0 to 10. The value 0 means: "not at all willing to take risks" and the value 10 means: "very willing to take risks". [scale 0 to 10]. *(For an analysis of the answers given, see Appendix A.6.)*

8. Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with other people? Please select a number on a scale from 0 to 10. The value 0 means: "can't be too careful" and the value 10 means: "most people can be trusted". [Scale 0 to 10] *(For an analysis of the answers given, see Appendix A.6.)*

9. What is your gender? *(For an analysis of the answers given, see Section 5.)*

10. What is your age [in years]? [*Integer*] *(For an analysis of the answers given, see Section 5.)*